THE ROLE OF ALTERNATIVES IN LANGUAGE

EDITED BY: Sophie Repp and Katharina Spalek
PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-840-3
DOI 10.3389/978-2-88966-840-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

THE ROLE OF ALTERNATIVES IN LANGUAGE

Topic Editors:
Sophie Repp, University of Cologne, Germany
Katharina Spalek, Humboldt University of Berlin, Germany

Citation: Repp, S., Spalek, K., eds. (2021). The Role of Alternatives in Language. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-840-3

Table of Contents

frontiers in Communication

REVIEW
published: 11 June 2021
doi: 10.3389/fcomm.2021.682009

The Role of Alternatives in Language

Sophie Repp[1]* *and Katharina Spalek*[2]

[1]Department of German Language and Literature I, University of Cologne, Cologne, Germany, [2]Department of German Language and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany

In this review we provide a discussion of the concept of alternatives and its role in linguistic and psycholinguistic theorizing in the context of the contributions that have appeared in the Frontiers Research Topic *The Role of Alternatives in Language*. We are discussing the linguistic phenomena for which alternatives have been argued to play a paramount role: negation, counterfactual sentences, scalar implicatures and exhaustivity, focus, contrastive topics, and sentences with bare plurals and with definite plurals. We review in how far alternatives are relevant for these phenomena and how this relevance has been captured by theoretical linguistic accounts. Regarding processing, we discuss the mental activation of alternatives: its mandatory vs. optional nature, its time course. We also address the methodological issue of how experimental studies operationalize alternatives. Finally, we explore the phenomenon of individual variation, which increasingly attracts attention in linguistics. In sum, this review gives an inclusive and broad discussion of alternatives by bringing together different research strands whose findings and theoretical proposals can advance our knowledge of alternatives in inspiring cross-fertilization.

Keywords: alternatives, negation, counterfactual, focus, scalar implicature, homogeneity, individual variation, alternative activation

OPEN ACCESS

Edited by:
Niels O. Schiller, Leiden University, Netherlands

Reviewed by:
Mante Sjouke Nieuwland, Max Planck Institute for Psycholinguistics, Netherlands
Anna Czypionka, University of Konstanz, Germany

Correspondence:
Sophie Repp
sophie.repp@uni-koeln.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *17 March 2021*
Accepted: *25 May 2021*
Published: *11 June 2021*

Citation:
Repp S and Spalek K (2021) The Role of Alternatives in Language. Front. Commun. 6:682009. doi: 10.3389/fcomm.2021.682009

INTRODUCTION

Many linguistic utterances convey meaning that must be interpreted against an alternative meaning in order to be fully informative, or even interpretable. For instance, when I say *If I were rich I would travel the world in 80 days*, I am talking about worlds where the proposition *I am rich* is true: I am (usually) saying that these worlds are non-factual worlds and that in the actual world the proposition *I am rich* is false. Thus, I am juxtaposing different (sets of) worlds. Another example is negation. When I say *Chris doesn't eat rhubarb* I am making an assertion about the actual world: in the actual world, the proposition *Chris doesn't eat rhubarb* is true. Usually, though, when I use negation, I am also considering (an) alternative (set of) world(s). In the present case these worlds would be worlds where the proposition *Chris eats rhubarb* is true. Again, we juxtapose alternative worlds.

Since propositions denote sets of worlds, another way of describing the above phenomena is to say that alternative propositions are involved in the interpretation of these sentences, for instance *{Chris eats rhubarb, Chris doesn't eat rhubarb}*. And since sentences are the linguistic objects that may denote propositions, substituting a linguistic expression in a sentence by a different expression will usually produce an alternative proposition. Consider *Chris ate some of the biscuits*. When we substitute *some* for *all*, the sentence changes its meaning: *Chris ate all of the biscuits* is true in different worlds. So the two sentences denote alternative propositions. The two expressions *some* and *all* also are alternatives—on the level of expressions, and on the level of the semantic objects that are denoted by these expressions, here quantificational determiners. Importantly, the presence of an expression like *some* typically in itself evokes alternatives: The sentence *Chris ate some of the biscuits* has the pragmatic meaning *Chris ate some but not all of the biscuits*, even though the denotation of *some* essentially is *some and possibly all* (i.e. *at least some*). This is a well-known scalar implicature: The use of an expression which has scalar alternatives—*some* and *all* are elements on a scale like *{none, some, many, most, all}* —, is interpreted as expressing the exclusion of the alternatives that lead to a stronger meaning[1] of the sentence, that is *{many, most, all}*. The alternative *none* is not directly relevant here because it is incompatible with the truth-conditional meaning of *some*. Thus, the sentence *Chris ate some of the biscuits* can only be interpreted in the intended way if the scalar alternatives of *some* are considered.

Alternatives also play a role in the information structuring of a clause, most notably for focus and for contrastive topics. Since Rooth (1985), Rooth (1992), and Krifka (2008) the alternative semantics view of focus has been very successful in the explanation of linguistic phenomena as well as in the description of language processing. According to this view, focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions (Krifka, 2008: 247). Focus interpretation is anaphoric in that the linguistic context must provide alternatives, or they must be easy to accommodate from the situational context. Consider the sentence *CHRIS ate the biscuits*, where the small capitals indicate prosodic prominence of the word *Chris*, i.e. the presence of a prominent pitch accent. Since one way of indicating focus in English is placing prosodic prominence on a syllable within the focus expression, ⟦Chris⟧ is the focus in this sentence. As a consequence, our sentence presupposes that ⟦Chris⟧ is an element in a non-singleton set of alternatives, where at least one other alternative is provided by the context, for instance *{Chris, Sam}*. Usually, our sentence will be taken to mean that substituting the focus by an alternative results in a false proposition: It is not true that Sam ate the biscuits. Thus, similarly to the case of the scalar implicature, the alternatives which are not the focus themselves are excluded. There are other instances of focus, where there is no exclusion of alternatives, e.g. when focus appears in the scope of additive focus-sensitive particles like *also* as in *Chris also ate* $\textit{CAKE}_{focus}$.

It is a natural question to ask at this stage if and how scalar alternatives and focus alternatives 'interact', and what the consequences are for the resulting meaning. How is focus on a scalar expression interpreted? Consider (1). In (1a), the accent takes the default position for sentences with wide focus on the whole sentence. In (1b), the accent indicates focus on the scalar expression *some* in the subject of the clause. So there are two sources for the relevance of alternatives in the interpretation: the lexical semantics of *some*, and the focus.

(1) a. Some of the kids ate CAKE.
 b. $[\text{SOME}]_{focus}$ of the kids ate cake.

The alternatives that are triggered by these sources arguably are the same, namely a set of expressions and their denoted meanings, e.g. *{none, some, many, most, all}*. Indeed, Fox and Katzir (2011) argue that focus alternatives and scalar alternatives are the same. As we discussed above, (1a) means *Some but not all of the kids ate cake* because the stronger alternatives are excluded. We also saw that without an additive focus-sensitive operator, focus is interpreted as signaling the exclusion of alternatives. What then is the meaning contribution of focus in (1b)?

To answer this question, it is important to remember that scalar implicatures do not arise obligatorily. In some contexts, scalar items receive the semantic *at least* reading. However, it has been found that placing an accent on a scalar item, i.e. focusing it, interacts with the computation of the implicature (e.g. Fretheim, 1992; Chevallier et al., 2008; Schwarz et al., 2008; Franke et al., 2017). Experimental results indicate that focus leads to a higher rate of implicature readings. These findings can be explained on the alternative semantics view that focus presupposes the existence of alternatives in the context. For (1b) this means that at least one of the alternatives of *some, {none, some, many, most, all}*, must have been mentioned in the discourse or must be easy to accommodate. Focus then leads to a 'reliable' exclusion of the contextual alternative. Note in this connection that scalarity is not restricted to quantificational elements. Many properties can be ordered on a scale, for instance *{cold, cool, warm, hot}*. Effects of prosody and thus focus on implicature computation were also suggested for these elements (e.g. Horn, 2006). Conversely, focus can easily evoke alternative sets involving a scale without an exclusion interpretation: It has long been noted that in sentences like *Chris only won the* $[\textit{BRONZE}]_{focus}$ *medal*, the focus-sensitive operator *only* is not used to exclude the alternatives (*silver* and *gold*): one can only win one medal in a competition anyway. Rather, the sentence expresses that *bronze* was less than what had been expected or wished for (Jacobs, 1991).

These examples show that alternatives play subtly different roles in different linguistic domains and that different 'types' of alternatives may interact. The goal of this review article is to evaluate the concept of *alternatives* in the linguistic domains where alternatives have been suggested to play a fundamental role in the interpretation and structuring of language. These domains have not been pursued completely independently of each other. For instance, counterfactuals, scalar implicatures and also focus often are discussed in relation to negation. However, work that puts negation into the center of its attention, usually asks and answers rather different questions than does work on counterfactuals, implicatures or focus. Our aim is to explore the various ways in which alternatives play a role for the different domains and to identify core characteristics of the notion *alternative* as well as potential differences between the domains. As the research on alternatives is carried out in different areas for each of which there exists a vast research literature, we cannot give a fully exhaustive review. To nevertheless achieve our aim, we are a giving a review of a selection of the existing literature which contextualizes the original research contributions in the Research Topic *The Role of Alternatives in Language* (marked with [RT] in this paper), and we will refer the reader to more specific reviews on individual topics in the respective sections of this paper. Some linguistic phenomena which might be considered to involve alternatives are beyond the scope of the current

[1]See Section *Negation and Counterfactuals* for elaboration on this notion.

review. These include for instance syntactic ambiguity, because we focus on alternatives from a semantic-pragmatic point of view, but also lexical ambiguity, which we do not consider for reasons of space.

We discuss the concept of alternatives both from the perspective of linguistic theory and from the perspective of language processing. Both fields have shown increased interest in alternatives in the last decades but there is not as much interchange as one would wish for. In our view, it is crucial to bring these fields together because the observation that alternatives are relevant or necessary for interpretation begs the question how these alternatives become part of the linguistic representation, which also is a *mental* representation. We may ask how and when alternatives are activated, and for how long they remain available—that is, what the temporal characteristics of alternative activation are, to what extent the exclusion of alternatives is an active mental process, and whether it requires mental resources. Issues like these might be relevant when it comes to grammatical conventionalization or to interpretation preferences, and the answers might be different for different types of alternatives.

The paper is structured as follows. In the section *Alternatives in Different Linguistic Domains* we discuss the linguistic domains where alternatives play a role by summarizing the core questions that have been asked for the respective domain, as well as the answers that have been proposed both on the basis of theoretical reasoning and on the basis of experimental evidence gathered with different methods. The discussion in this section addresses characteristics of the representation of alternatives in these domains both from a linguistic (grammatical) point of view and from a mental model (psychological) point of view. The section *The Activation Process* focuses on the issues of processing listed above, that is on characteristics of the *process* of constructing the representation of alternatives, for instance temporal aspects. In the section *Alternatives in the Lab*, we investigate how alternatives as linguistic objects can be related to possible worlds in experimental settings, that is we discuss the operationalization of alternatives, which is an important issue for the experimental investigation of both the grammar and the processing of phenomena involving alternatives. The section *Alternatives for All? Individual Differences* addresses an issue that has become increasingly relevant in recent years also both in linguistics and in psycho-/neurolinguistics: the extent and evaluation of individual differences. These are pervasive in the realm of alternatives, too, and pose interesting questions for linguistic and psycholinguistic theories. The section *Conclusion* summarizes and concludes.

ALTERNATIVES IN DIFFERENT LINGUISTIC DOMAINS

Negation and Counterfactuals

As mentioned in the *Introduction*, for negation the alternative propositions p and $\neg p$ are relevant: $\neg p$ is what a negative sentence denotes, p is negated. As simple as it sounds, this state of affairs has non-trivial consequences for the factors that influence the grammar and processing of negation. We will focus on two factors in this section: semantic and world knowledge on the one hand, and discourse context on the other hand.

Semantic and World Knowledge and Negation Alternatives

Negative sentences normally are used to express the falsity of a proposition p whose truth or falsity is at issue. This means that there must be a chance of p actually being true. A sentence like *A robin is not a tree*, although true, is unlikely to be uttered because the negated proposition p (*A robin is a tree*) is unlikely to be true outside very specific contexts. Therefore, sentences negating an unlikely proposition p are usually pragmatically infelicitous and it has been shown that they incur increased processing costs and are harder to recall than pragmatically felicitous negative sentences (e.g., Wason, 1965; Cornish, 1971; Hörmann, 1971; Johnson-Laird and Tridgell, 1972; Wason, 1972; Givon, 1978; Watson, 1979; Arroyo, 1982; Glenberg et al., 1999; Nieuwland and Kuperberg, 2008; Orenes et al., 2016).

The pragmatic (in)felicity of negative sentences is fed by semantic and world knowledge: We know that robins and trees are living organisms in different biological kingdoms—animals and plants—and it is not informative to say that a particular genus of animal is not a clade in the plant kingdom. A true positive proposition like *A robin is a bird*, in contrast, can be considered informative if the addressee does not know what particular animal a robin is. There have been numerous studies investigating the role of pragmatic felicity in relation to semantic and world knowledge, amongst them two studies that have appeared in the Research Topic *The Role of Alternatives in Language* (2019). Haase et al. (2019)[RT] investigated pragmatic plausibility via the co-hyponym relation such that there is a large semantic feature overlap between alternatives (unlike for birds and trees), and both alternatives in principle would be pragmatically plausible in a sentence such as *George Clooney is (not) an actor/singer*. Haase et al. employed the method of event-related brain potentials, which in earlier research yielded the result that for stimulus sets without pragmatic control the N400 component on the final word is larger for false than for true affirmative sentences, whereas for negative sentences it is the other way round (e.g., Fischler et al., 1983; Kounios and Holcomb, 1992; Lüedtke et al., 2008; Nieuwland and Kuperberg, 2008; Wiswede et al., 2013; Dudschig et al., 2016; for early research using behavioral methods reporting this truth-polarity interaction, see e.g., Wason and Jones, 1963; Gough, 1965; Clark and Chase, 1972; Carpenter and Just, 1975; for a recent review also on findings not showing the interaction, see Kaup and Dudschig, 2020). The truth-polarity interaction has been suggested to result from the semantic subject-predicate mismatch (e.g., *robin—tree*) in the false affirmative and the true negative conditions, and has sparked discussion about a late integration of the meaning of the negative marker, to which we turn in the section *The Activation Process*. As for the role of pragmatic felicity, Haase et al. replicate the above-mentioned ERP-findings for the affirmative sentences but for the negative sentences the (reverse) effect is non-significant. The authors argue that when the correctly negated predicate is a co-hyponym and thus pragmatically felicitous, anticipation of the negated material is more successful than when there is no co-hyponymy (they do not take a stance as to whether the N400 reflects expectancy, prediction or integration). Their finding supports the relevance of pragmatic felicity for the interpretation of negative sentences, adding to the previous findings on this issue.

Pragmatics as fed by semantic and world knowledge can also play a role via conventionalization (Kronmüller and Barr, 2015; Kronmüller et al., 2017). Kronmüller and Noveck (2019)[RT] present evidence from a negative reference task where participants picked objects with or without conventional names. A conventional name is for instance *vase* for a vase. An object without a conventional name is for instance an unusually shaped clay form. Participants were instructed not to pick for instance *the sculpture*, where *the sculpture* refers to an unusual object that in the course of the task had been *ad-hoc*-named *the sculpture*. In addition to "the sculpture," there were two alternative objects. When instructed not to pick "the sculpture," the choice of object by the participants depended on whether the other objects both were unusual and on whether they had a conventional or *ad-hoc* name (given in the course of the task) or not. The authors show that linguistic conventions (names) and contextual linguistic and visual information influence the inferences listeners make about alternatives when interpreting negative sentences.

The Discourse Context

A proposition $\neg p$ is true in many different (possible) worlds: saying what is not the case is not yet saying what is the case. Therefore, the question arises what representation we form when we hear a negative sentence. From a grammatical point of view, we would say that it is a representation containing a negation operator associated with p, so that two propositions (p and $p' = \neg p$) are elements of the linguistic representation. From a mental model point of view, the representation might also be one of a proposition q, which is true in a subset of the $\neg p$-worlds but tells us something more about those worlds. For example, when we hear *Chris didn't eat cake*, we might form a mental model corresponding to *Chris ate biscuits*, because we know that biscuits were under consideration in the context (unless we think that Chris didn't eat anything at all). Thus, the discourse context is important for negation because it typically restricts the number and type of relevant alternatives when interpreting a negative sentence. This is true especially for the context preceding the negative sentence, but we will see further below that the context following a negative sentence also is important: it influences the anaphoric uptake of the propositional alternatives p and $p' = \neg p$.

One area where the preceding discourse context has been found to be important is the above-mentioned truth-polarity interaction, which early on has been suggested to result from a two-step construction of the representation of negative sentences (Carpenter and Just, 1975). In recent theorizing, this two-step construction is formulated as two-step simulation of the mental model representing the situation described by the negative sentence (Kaup and Zwaan, 2003; Kaup et al., 2006). The claim is that first, a situation model is simulated which corresponds to the (false) positive proposition p—the *affirmative* situation—and then a model of the actual, *negative* situation is simulated. Two-step simulation has been shown with different methodologies (e.g., Kaup et al., 2006; Kaup et al., 2007; Dale and Duran, 2011; Autry and Levine, 2012; Orenes et al., 2014; Orenes et al., 2016) but there are also proposals that negation is processed in one step (e.g., Mayo et al., 2004; Anderson et al., 2010; Papeo et al., 2016). For recent reviews on these issues, see Tian and Breheny (2019) and Kaup and Dudschig (2020). Whether or not a simulation happens in two steps depends (inter alia) on the number and kind of alternatives that are available in the discourse. In the following, we will discuss properties of the preceding discourse context that are relevant with respect to this issue. The actual time course of the activation of negation alternatives will be discussed in *The Activation Process*.

Consider the sentence *The window is not open* with the predicate *open*, which has a contradictory antonym: *closed*. It is easy to simulate a situation model of the actual, negative situation because that model must have a closed window in it. So there are two alternatives available that may be simulated in two steps. For predicates with multiple, i.e. contrary antonyms (e.g., *green*), it is unclear what the negative situation looks like because there are many options for the second alternative: blue, red etc. The linguistic or situational context may restrict the number of such alternatives and thus provide 'more suitable' content for a simulation of the negative situation, and it has been shown that this has an impact on negation processing (e.g., Wason, 1961; Kroll and Corrigan, 1981, also see Mayo et al., 2004). We are illustrating the relevance of the number of alternatives here with findings from Orenes et al. (2014), who conducted a multi-picture visual-world paradigm with four different colors where a context sentence announced either two alternative colors (e.g., *green*, *blue*), or more than two alternative colors (e.g., *green*, *blue*, *yellow*, *pink*) to be present in the picture. They found that in the two-alternative context, participants listening to negative sentences like *The figure is not green* briefly looked at the green figure and then focused their attention on the blue figure, i.e. the figure representing the one contextually plausible negative situation. In the multiple-alternative context, participants focused on the green figure only. Thus, when the context restricted the alternatives to two, two alternatives are activated, otherwise this is not the case.

The preceding context may also influence the availability of the affirmative situation by providing several $\neg q$-alternatives $\{\neg q_1, \neg q_2, \ldots\}$. Assume we wish to ask some students for their opinions about a city they have never been to. To establish the set of appropriate addressees, we ask the students: *Who hasn't been to Bielefeld?* This gives us the set of students who have not been to Bielefeld, and by extension a set of alternative negative propositions, e.g., *{Alex hasn't been to Bielefeld, Chris hasn't been to Bielefeld, Robin hasn't been to Bielefeld}*. This set does not contain a positive proposition. For negative sentences uttered in the context of such a question, there does not seem to be two-step simulation (Tian et al., 2010; Tian and Breheny, 2016). The context question is the so-called *Question under Discussion* QuD

(Roberts, 1996; cf. von Stutterheim, 1997). It introduces and restricts focus alternatives: The answer to the QuD must contain an alternative that is an element in the focus alternative set introduced by the QuD. Thus, similarly to the interaction of focus alternatives and scalar alternatives (see *Introduction*), we find that the mental representation and linguistic relevance of alternatives from different domains interact. The alternative set relevant for negation is *{p, ¬p}*, but with a QuD like the above, the *p*-alternative is not relevant/salient, whereas the focus alternatives prescribed by the QuD are. The interplay of focus and negation has received considerable interest in the theoretical linguistic literature (e.g., Jackendoff, 1972; Jacobs, 1991; Partee, 1993; Hajičová, 1996; Rooth, 1996; Herburger, 2000; Beaver and Clark, 2008; Büring, 2016b). For a recent review, see Fălăuş (2020).

The relevance of the subsequent discourse for negation alternatives has only recently attracted attention. As already mentioned, in linguistic theorizing it is typically assumed that a negative sentence introduces two propositions that are principally available for anaphoric reference: *p* and *¬p* (e.g., Krifka, 2013; Meijer, 2016; Meijer and Repp, 2018; Claus et al., 2019). Which proposition gets picked up by a propositional anaphor depends on characteristics of the subsequent discourse. Experimental evidence from acceptability judgment studies suggests that propositional anaphors like *that* preferentially are interpreted as taking the negative proposition *¬p* as antecedent (Claus et al., 2019). *¬p* is of course the proposition that is denoted by the negative sentence (and that is associated with the final situation model). Therefore, a general preference for *¬p* might not be surprising. Importantly, certain parameters in the sentence containing the anaphor might change this preference. For instance, modal particles, belief-state verbs vs. reporting verbs, and counterfactual tenses have all been suggested to play a role. Claus et al. (2019) show for German that the presence of the adversative conjunction *but* vs. the focus particle *auch* ('also/too') in a dialogue like the following changes the preference from *¬p* to *p*: *A: Tom didn't steal the bag. B: Jenny believes that*$_{\neg p}$*, too./But Jenny believes that*$_{p}$. Meijer and Repp (2018), also investigating German, find a shift to *p*-interpretations that is triggered by tense and by a modal particle. Illustrating the former, when someone says *Alex wasn't here on Monday*, a response like *That would be weird* typically is interpreted as meaning that Alex's absence would be weird, i.e. *that* refers to *¬p*. A response like *That would have been weird*, in contrast, is typically interpreted as meaning that Alex's presence would have been weird., i.e. *that* refers to *p*. The authors argue that the response in these examples is the consequent of a contextual counterfactual and means *It would have been weird if Alex had worked on Monday*. The interpretive difference between the tenses results from the specific way tense is interpreted in the consequent of counterfactuals. The findings from this literature suggest that negative sentences indeed make available a positive proposition that principally is available for anaphoric uptake.

Staying briefly with counterfactual sentences, we note that they are generally interesting for negation alternatives because (typically) a positive sentence describes a negative situation—and the factual and a non-factual world are juxtaposed. In indicative conditionals, the alternatives *{p,¬p}* also are relevant. Indicatives differ from counterfactuals in the tense of antecedent and consequent (counterfactual—past: *if there had been... there would have been...*; indicative—present: *if there are... there are...*). In an indicative, an antecedent containing the proposition *p* restricts the truth of the consequent proposition *q* to *p*-worlds and excludes *¬p*-worlds. In a counterfactual, the alternativeness is intuitively more prominent because the actual world (usually) is assumed to be false and the non-factual worlds are the worlds 'of interest.' In the theoretical literature, it is usually assumed that for a counterfactual to be plausible the factual and the non-factual worlds must be very similar (Goodman, 1955; Stalnaker, 1968; Lewis, 1973). We cannot address this issue here for reasons of space but we would like to point out that the similarity approach has recently been argued to be problematic (Ciardelli et al., 2018). The empirical argument involves *if*-clauses with negation, which seem to be judged differently than would be expected by the similarity approach. However, Schulz (2019) argues that the negation in the *if*-clause introduces alternatives which become relevant for the interpretation, similarly to alternatives that are introduced by a disjunction in the *if*-clause (e.g., Alonso-Ovalle, 2009; Santorio, 2018; Willer, 2018). Thus, we observe that there are intricate interactions between different 'types' of alternatives in the domain of counterfactual conditionals, too. For a recent review on counterfactuals, see Arregui (2020). We return to counterfactual and indicative conditional sentences in the section *The Activation Process*, when we discuss issues of processing.

(Scalar) Implicatures and Exhaustivity

In the *Introduction* we discussed scalar implicatures involving scales for quantificational determiners like *some*, and scales for predicates like *warm*. These scales were scalar in an intuitive sense. A scale like *{none, some, many, most, all}* is a scale of quantities, which is something we can measure in the real world even if the threshold for using a particular element on the scale rather than its neighbor is not clear in every case. Similarly, scales like *{cold, cool, warm, hot}* concern properties in the real world. In this section, we discuss implicatures that might be considered to be less intuitively scalar because the relation of strength, which we appealed to when we talked about the exclusion of scalar alternatives, seems to concern information states or knowledge about the world—or indeed, the common ground –, and not objects or properties in the world.

The Gricean and the Grammatical View of Implicatures

Assume a context where two people are under consideration for having been invited: Chris and Alex. A dialogue ensues: A: *Who did Toni invite?* B: *Chris*. B's answer is usually taken to mean that B invited Chris but not Alex. This is a run-of-the-mill Gricean implicature (Grice, 1967), where the meaning of B's answer is a *strengthened meaning*. The traditional, Gricean explanation for why this meaning arises is the following (see e.g., Horn, 1972; Grice, 1975; Gazdar, 1979; Soames, 1982; Levinson, 2000). Assuming that B is cooperative, their answer will be relevant in the context of the question, and it will entail all the information that is compatible with B's knowledge. If B knew that Toni invited Chris and Alex, saying so would have conveyed a stronger meaning: the proposition *Toni invited Chris and Alex* is true in fewer worlds than the proposition *Toni invited Chris* because the latter is also true in worlds where Toni invited Chris but not Alex. Thus, a stronger meaning is one where we know more about the world we are in. Since B did not choose to express the stronger meaning, B's answer implicates that the stronger meaning is not true—it is excluded—and we arrive at the strengthened meaning *Toni invited Chris but not Alex*.

It has long been noted that the original Gricean view faces problems, which have fueled the development of two different types of theoretical accounts. One of the most-discussed phenomena in this regard are embedded implicatures, which arise in the scope of a higher operator. Returning to our example from the *Introduction*, *Some of the kids ate cake*, recall that the strengthened meaning is *Some but not all of the kids ate cake.* Embedding our example sentence under the verb *know* yields: *Toni knows that some of the kids ate cake*. Intuitively, the sentence means *Toni knows that some of the kids ate cake and Toni knows that not all of the kids ate cake.* It does not mean *Toni knows that some of the kids ate cake and it is not the case that Toni knows that all of the kids ate cake*, which, however, would be expected under a Gricean account. One type of account assumes that implicatures of this sort arise due to a silent exhaustification operator, *EXH* (alternatively O_{alt}/O_c)). EXH can be introduced locally, but is (also) introduced by default at the top node of every matrix sentence. EXH exhausts the alternatives, i.e. it excludes stronger alternatives, so that a strengthened meaning arises. This view, which derives implicatures involving exhaustivity effects compositionally, is the *grammatical* view of implicatures and was first proposed by Chierchia (2004), Chierchia (2006); also see Katzir (2007), Fox and Katzir (2011), Chierchia et al. (2012), Trinh and Haida (2015). The other type of account are the (Neo-)Gricean approaches, including rationalist accounts and relevance-based accounts, which place pragmatic reasoning at the center of the explanatory framework (e.g., Sauerland 2004; Van Rooij and Schulz, 2004; Schulz and Van Rooij, 2006; Benz and Van Rooij, 2007; Geurts, 2009; Geurts, 2010; Franke, 2011; Russell, 2012; Frank and Goodman, 2012; Frank and Goodman, 2014; also see Benz and Stevens, 2018). There are also accounts explicitly combining aspects of both theories (e.g., Potts et al., 2015). For recent reviews on theories of implicatures, see for instance Breheny (2019) or Nicolae and Sauerland (2020).

Characteristics of Scalar Alternatives

There are several contributions in the Research Topic *The Role of Alternatives in Language* that explore specific characteristics of scalar alternatives. Trinh (2019)[RT] investigates the contextual source of alternatives that are relevant for implicature computation. Tomioka (2021)[RT] and Singh (2019)[RT] explore implicatures in different subdomains—disjunction, quantificational determiners, numerals and so-called free-choice implicatures. Feng and Cho (2019)[RT] consider so-called *indirect implicatures*. In contrast to the *direct implicatures* we have investigated so far, indirect implicatures arise when a scalar term at the *endpoint* of a scale is negated: *Toni did not always go to the beach last week* implicates that it is not the case that Toni never went to the beach last week, that is sometimes, Toni did go to the beach. Both Singh (2019)[RT] and Feng and Cho (2019)[RT] investigate the processing cost of these various implicatures. We will discuss their contributions in the section *The Activation Process*, where we will see that the computation of strengthened meanings can but need not be costly (e.g., Bott and Noveck, 2004; Breheny et al., 2006; Huang and Snedeker, 2009; Marty et al., 2013; Chemla and Bott, 2014; Cremers and Chemla, 2014; Benz and Gotzner, 2020). In this section we concentrate on the role of the context and discourse factors more generally.

Trinh (2019)[RT] explores how the set of alternatives that are relevant in the computation of exhaustivity implicatures can be restricted in a way that predicts intuitively correct implicatures in a number of subdomains. For instance, our question-answer discourse above might have given the impression that the question provides the relevant alternative set—as it would do in the case of focus alternatives: *{Toni invited Chris, Toni invited Alex}*. However, recall that it is actually the conjunction of these propositions that is the relevant stronger alternative for the exhaustivity implicature of the response. So the alternative set should be *{Toni asked Chris, Toni asked Alex, Toni asked Chris and Toni asked Alex}*. But then, why should we not also make *Alex did not ask Chris* part of the alternative set? After all, we are interested in who Toni asked—and thus also in whom they did not ask. Trinh discusses issues like these within the grammatical approach of implicatures and explores three notions that may be used to restrict the alternative set: relevance (closure under Boolean operations), utterance (what was explicitly uttered in the linguistic context: "formal alternatives") and salience (what is contextually salient). Trinh shows that these notions make distinct predictions for the computation of implicatures and argues that salience poses some non-trivial problems.

Tomioka (2021)[RT] investigates disjunction within the grammatical approach to implicatures. In the exhaustivity literature, disjunction has been argued to display some special characteristics. To illustrate, Hurford's constraint (Hurford, 1974) says that one disjunct must not entail the other: #*Toni traveled to Cologne or to Germany* is infelicitous, because traveling to Cologne entails traveling to Germany. Interestingly, this constraint does not apply to disjunctions of scalar alternatives (Gazdar, 1979): *Toni read some or all of the books*. This observation has been explained in the grammatical approach by assuming that the EXH operator applies to the first disjunct before the disjunction is interpreted. EXH changes the meaning of the first disjunct and thus removes the entailment relation: *Toni read some but not all of the books, or all of the books*. The order of the disjuncts is relevant here (Singh, 2008): #*Toni read all or some of the books* is infelicitous without a lexical *only* in the second disjunct. This observation is explained as a result of incremental (i.e. left-to-right) computation of the disjuncts, which makes insertion of EXH into the second disjunct ineffectual, which is disallowed (Singh, 2008; Fox and Spector, 2018).

Tomioka (2021)[RT] shows that the empirical observations about disjunctions carry over to coordinations, subordinations and even independent sentences that are separated by a turn-take in conversation. Compare the following contrastive coordinations: #*Toni traveled to Cologne but Alex to Germany*; *Toni read some of the books but Alex read all of them*; #*Toni read all of the books but Alex read* #*(only) some of them*. Tomioka argues that the previous accounts cannot explain these facts because they rely on disjunction, and proposes that the alternatives that are at issue here, are focus alternatives: *Toni read [some]$_{focus}$ of the books or Toni read [all]$_{focus}$ of the books*. He proposes the *Contrast Antecedent Condition*, which requires that the first of the two sentences provides an antecedent from which a focus alternative set can be generated that fits the focus in the second sentence. Furthermore, the focus alternative set must contain mutually exclusive alternatives comprising the meaning of both sentences. The asymmetry of Hurford's constraint is explained (roughly) as follows: If the meaning of the first sentence can be strengthened (e.g., *some... → some and not all...*) and the meaning of the second sentence (e.g., *all...*) provides a mutually exclusive alternative, the result is felicitous. If the meaning of the first sentence cannot be strengthened (e.g., *all... →/*), the meaning of the second sentence (e.g., *some...*) does not provide the required alternative. By appealing to focus rather than scalar alternatives, Tomioka explains the observed effects in structures beyond disjunctions.

Focus and Contrastive Topics

In the previous sections, we repeatedly encountered focus alternatives. We saw that focus interacts with other types of alternatives (negation, scalars), and that focus alternatives for some phenomena might offer the better explanation than other alternatives (scalars). In this section, we consider research where focus alternatives are the central object of investigation. One strand of this research is concerned with the prosodic marking of focus. For instance, for intonation languages it has been shown that certain accents trigger the activation of alternatives during comprehension. Often, arguably more prominent accents, like English L+H* in contrast to H* lead to a more reliable activation of alternatives (e.g., Dahan et al., 2002; Weber et al., 2006; Ito and Speer, 2008; Watson et al., 2008; Braun and Tagliapietra, 2010; Husband and Ferreira, 2016; Braun et al., 2018). Yan and Calhoun (2019)[RT] show for a language which marks focus prosodically not through accenting but through pitch range extension—Mandarin—that this kind of prosodic prominence also triggers focus alternatives. Furthermore, the choice of accentuation pattern in intonation languages has been shown to influence memory retrieval (e.g., Fraundorf et al., 2010; Gotzner et al., 2013; Repp and Drenhaus, 2015; Gotzner, 2017; Tjuka et al., 2020). Finally, the presence of elements that require focus for semantic reasons, like the focus particle *only*, influences the processing and memory of accented words and their alternatives (Spalek et al., 2014; Gotzner et al., 2016). See section *The Activation Process* for more details on processing.

Focus alternatives can be triggered by prosodic means without discourse context. Since focus is anaphoric, this begs the question what exactly serves as a suitable focus alternative. Building on some of the studies mentioned above, Braun and Biezma (2019)[RT], and Yan and Calhoun (2019)[RT] compared different types of potential focus alternatives: what they call *contrastive alternatives* (words from the same semantic field as the focused expression, e.g., *swimmer—diver*), and *non-contrastive* alternatives (words which are semantically related via the event, e.g., *swimmer—pool*). Braun and Biezma consider the activation of alternatives by a prenuclear accent that arguably marks a contrastive topic in German (L*+H). Contrastive topics have received various analyses, all of which involve information-structural alternatives: either 'simple' focus alternatives, or more complex alternatives, which may also reflect a specific QuD-induced discourse (e.g., Büring, 2003; Wagner 2012; Constant, 2014; Büring, 2016a). Braun and Biezma report that prenuclear L*+H activates alternatives quite similarly as a nuclear focus accent does, suggesting that a parsimonious analysis of contrastive topics should assume contrastive topics to be as similar to focus as possible.

The issue of what a suitable focus alternative is does not only arise in the absence of context. For the additive particle *also*, this issue is notorious. It has been observed that for sentences with *also* a relevant alternative must have been uttered in the context or be entailed by it. Accommodation usually fails, different from other focus particles: *Tim is a sugar addict. He even/#also eats [candied FLIES]$_{focus}$*. Recall that similarly subtle restrictions on the contextual availability of alternatives have been discussed for scalar alternatives by Trinh (2019)[RT] (section *(Scalar) Implicatures and Exhaustivity*). The controversy surrounding *also* is whether the context must provide an alternative proposition (e.g., *He eats marshmallows*) or whether an alternative to the focused constituent is sufficient (*marshmallows*) (Corblin, 1991; Heim, 1992; Asher and Lascarides, 1998; Geurts and van der Sandt, 2004; Roberts, 2010; Tonhauser et al., 2013; Ruys, 2015). Grubic and Wierzba (2019)[RT] discuss this issue for the German particle *auch* 'too.' On the basis of experimental evidence, they argue that positing the presence of a propositional alternative is too strict but a sub-propositional alternative neither is sufficient to make the use of *auch* felicitous. However, it can be sufficient if merely accommodating the propositional alternative makes the discourse more coherent. We will see in the section *The Activation Process* that for scalar alternatives, comparable suggestions have been made (Singh, 2019[RT]): Creating coherence or/and answering a QuD more completely are objectives that influence the cost and success of the computation of the meaning of alternatives.

For some phenomena involving focus alternatives, focus itself is not enough to create coherence. Cleft sentences are an example. They are well-known to show exhaustivity effects, for which it is under debate whether they are a presupposition or an implicature (e.g., Halvorsen, 1978; Atlas and Levinson, 1981; Horn, 1981; Percus, 1997; Velleman et al., 2012; Büring and Križ, 2013; De Veaugh-Geiss et al., 2018; see e.g., Onea 2019 for a recent review). Furthermore, clefts often are thought to involve contrastive focus, where the notion of contrast is somewhat unspecified (Repp, 2016 for a discussion of *contrast*). Destruel et al. (2019)[RT] argue for English and French, that the contrast in clefts is a doxastic type of contrast concerning the interlocutors' expectations (*contrariness* in Zimmermann (2008)): A cleft signals a stark contrast between what has been said or insinuated by another person and what the speaker assumes, that is the focus alternatives are restricted by the discourse context and the contrast concerns properties of the discourse.

This, or a similar discourse property of clefts might also be responsible for the observation in Yan and Calhoun (2019)[RT] that syntactic focus marking in Mandarin through clefting does not activate alternatives (whereas prosodic marking does). The authors suspect that clefts in Mandarin might require different context conditions from 'plain' focus, which in their experiments were not given. They also speculate that prosodic prominence and not focus might trigger alternatives, which would be compatible with Braun and Biezma's (2019)[RT] assumption that contrastive topics and focus are similar in terms of alternative activation. However, such a hypothesis leaves open what kind of alternatives the alternatives are from a semantic point of view: they do seem to be restricted to *contrastive* alternatives (*swimmer—diver*). Note, incidentally, that these are co-hyponyms, which are also relevant for negation alternatives (*Negation and Counterfactuals*).

Generics and Plural Definites

The last domain that we will discuss here are particular nominal expressions that occur in the subject position of a sentence: bare plurals as in *Beetles fly*, and definite plurals as in *The beetles are red*. The issue here roughly is to what extent the group of individuals denoted by these expressions must (not) be homogeneous in having the property expressed by the predicate of the sentence. Alternatives come into play in various ways.

The sentence with a bare plural, *Beetles fly*, is a generic sentence: It makes a generalization and is true although not all beetles fly—generalizations allow for exceptions. The term *exception* suggests that the predicate in a generic sentence must apply to the majority of the individuals of interest, contrary to fact: *Birds lay eggs* is felicitous although less than the majority of birds lay eggs—male birds do not. Conversely, there are generic sentences where the predicate does apply to the majority of individuals, and nevertheless they are infelicitous: *[?]Germans are right-handed* (cf. Carlson, 1977). Thus, an approach to generics which relies on statistical information like *a majority of x* and calculates the probability of having a certain property (Cohen, 1999; Cohen, 2004) seems problematic. However, the other prominent approach to generics, which is based on assumptions about *normal* individuals or circumstances (e.g., Asher and Morreau, 1995) quite clearly also faces problems.

In probabilistic approaches, alternatives have been used to keep the intuition about the majority rule: the set of individuals may be restricted by a set of alternatives to the predicate (Cohen, 1999). For instance, for *Birds lay eggs*, the set of birds intersects with a set of predicates which only apply to female animals, e.g., *{lay eggs, give life birth}*. Consequently, the relevant set for the majority rule is female birds. Including predicate alternatives this way yields the so-called *absolute reading* of generics sentences (Cohen, 1999): the sentence asserts something about birds *without* comparing birds to other individuals. Generics may also have a *relative reading*. For instance, *Dutchmen are good sailors* is felicitous because Dutchmen are compared to other nationalities, and not because the absolute majority of Dutchmen (even if restricted by a predicate alternative set) are good sailors. Cohen suggests that on the relative reading, a generic sentence must have an accent on the subject: to indicate the relevant focus alternatives. In the absolute reading, the predicate is in focus.

Kochari et al. (2020)[RT] explore subject and predicate alternatives and the associated readings in detail for English and propose that the relative reading is the basic meaning (cp. Tessler and Goodman, 2019): On the one hand, the relative reading reduces to the absolute reading if no alternatives for the subject are available. On the other hand, the relative reading finds grounding in learning mechanisms, which is critical for generics because making generalizations requires learning about the world. Kochari et al. also argue for a third type of alternative: causal background factors. These factors are additional properties that are causally relevant for the individuals having the property at issue in the generic sentence.

For sentences with plural definites, homogeneity becomes relevant as follows. *The beetles are red* seems to be truth-conditionally equivalent to *All the beetles are red* and express universal quantification. However, in a situation where some beetles are red and others are blue—i.e. a "non-homogeneous" situation –, the sentence with the definite is judged to be neither true nor false whereas the sentence with the universal quantifier is simply false. For the negated versions of the two sentences, there is the same discrepancy. This characteristic of plural definites—to be neither true nor false if the group of individuals is non-homogeneous in the property of interest—is known as *homogeneity* or *gappiness* effect, because there seems to be a truth value gap (Fodor, 1970; Löbner, 1987; Schwarzschild, 1994; Löbner, 2000; Magri, 2014; Križ, 2015). One explanation for the effect is that plural definites come with a maximality presupposition (e.g., Schwarzschild, 1994; Löbner, 2000). However, this clashes with the observation that in certain contexts, non-maximal readings are available. For instance, after a party a sentence like *The guests were happy* can be true even if not every single one of the guests was happy (Dowty, 1987; Malamud, 2012). Another explanation is that plural definites are semantically underdetermined, and may receive an existential or a universal reading, depending on the context (downward-entailing/upward-entailing, Krifka, 1996; /non-monotonic, Malamud, 2012). The existential and the universal reading are scalar alternatives, which can be the basis for scalar strengthening (Krifka, 1996; Magri, 2014; Križ 2015; Bar-Lev, 2018; Križ and Spector, 2020). In the sections *Alternatives in the Lab* and *Alternatives for All? Individual Differences* we come back to homogeneity effects of plural definites when we discuss the contribution by Tieu et al. (2019)[RT] in relation to the experimental operationalization of alternatives and individual variation.

This concludes our review of linguistic domains that have been discussed in relation to alternatives. We see that often it is not clear yet what kind of alternatives should be assumed to explain a certain meaning aspect, but also that it is highly plausible that there are different types of alternatives because they can be 'combined'—as is the case for focus and scalar alternatives. We also see that the range of factors determining the selection of (relevant) alternatives are manifold. Overall, some alternatives are more context-dependent than others but context and discourse coherence—maybe unsurprisingly—always play a role.

THE ACTIVATION PROCESS

This section is concerned with the *process* of activating alternatives. It addresses two questions: 1) whether the activation of alternatives is mandatory for the language processing of the phenomena discussed above or whether alternatives are activated strategically by the language users, and 2) what the time course of this activation process is.

The Nature of the Activation Process (Mandatory or Strategic)

The question whether the activation of alternatives is mandatory (automatic) or whether it underlies the strategic control of language users, has been explored for all the domains where alternatives are relevant that we discussed in the section *Alternatives in Different Linguistic Domains*. Starting with negation, we already mentioned that two-step theories posit the mandatory activation/simulation of the affirmative alternative before the negative alternative. In one-step theories, in contrast, the negative situation is available immediately. However, since most one-step theories assume inhibition of the affirmative situation, both alternatives are activated simultaneously and hence, again, activation is assumed to be mandatory. In the Research Topic *The Role of Alternatives in Language*, the contribution by Beltrán et al. (2019)[RT] provides evidence for the inhibition and thus mandatory activation of the affirmative alternative. The authors combined the comprehension of positive or negative action or non-action sentences with a go/no-go paradigm while measuring event-related potentials. They observed evidence for inhibition-related effects for negative sentences. Importantly, these effects were independent of the action/non-action sentence type, suggesting that negation triggers inhibition, which indicates mandatory activation of the positive alternative. However, evidence has been accumulating that the negative alternative can be accessed directly and without inhibitory effects if it is the most plausible one in the context (e.g., Nieuwland and Kuperberg, 2008; Dale and Duran, 2011; Autry and Levine, 2012). Hence, it has become a fruitful research endeavor to learn more about the circumstances under which both alternatives are activated. We discussed some of these circumstances in the section *Negation and Counterfactuals*.

Evidence from individual variation (see section *Alternatives for All? Individual Differences*) is also informative about the mandatory vs. strategic nature of the activation process for negation alternatives: In an eyetracking study on counterfactuals using pictorial displays reported in Orenes et al. (2019)[RT], participants listened to sentences like *If there had been oranges, there would have been pears*, having to infer that, in fact, there are no oranges and no pears. Initially, that is within about half a second, a significant group of participants increased their looks to both the real-world alternative and to the counterfactual alternative, suggesting parallel activation. Another group looked only at the real-world alternative. This finding may be taken to suggest that the activation of the alternative is not mandatory for the negation alternatives in counterfactuals. Kulakova and Nieuwland (2016a) review the literature on the processing of counterfactuals and conclude that while a dual linguistic representation of both p & q and $\neg p$ & $\neg q$ seems to be almost part of the definition of counterfactuals, convincing evidence for the synchronous availability of both representations is hard to come by. This conclusion indirectly supports our assumption that the findings by Orenes et al. (2019)[RT] speak against a mandatory process. However, as we will see in the section *Alternatives for All? Individual Differences*, there is another explanation available for these observations. Furthermore, other studies, for example a priming study by Santamaria et al. (2005), do support the assumption that both representations (i.e., p & q and $\neg p$ & $\neg q$) are (generally) available simultaneously (also see Thompson and Byrne, 2002; Byrne, 2005 for discussion).

Turning to scalar implicatures, recall from the *Introduction* that implicatures are not an obligatory part of the 'final' meaning of a sentence. Furthermore, we briefly mentioned in the section *(Scalar) Implicatures and Exhaustivity* that different types of scalar implicatures seem to come with different processing costs. Singh (2019)[RT] observes that for scales of quantificational determiners and of logical operators, the strengthened meaning seems to incur higher processing costs than the non-strengthened meaning; for scales of numerals and so-called free-choice implicatures, it is the other way round (e.g., Noveck and Posada, 2003; Bott and Noveck, 2004; Breheny et al., 2006; Chemla, 2009; Huang and Snedeker, 2009; Marty et al., 2013; Chemla and Bott, 2014; cp. Chemla and Singh, 2014; Crnič et al., 2015; Chemla et al., 2016; van Tiel and Schaeken, 2017; for discussion). These observations might be taken to suggest that for some implicatures the computation of the strengthened meaning of a sentence is an additional, non-mandatory process, but for other implicatures it is not. Singh (2019)[RT] suggests that rather than explaining the difference on the basis of the particular implicature computation, the two resulting meanings should be compared: 1) in relation to their semantic complexity (~presence of EXH), and 2) their usefulness in resolving uncertainties in a discourse (~to what extent they answer the QuD; formulated in terms of *entropy*; Shannon, 1948; van Rooij, 2004). Singh suggests that semantic complexity may increase cost for a meaning but if a more complex meaning helps reducing uncertainty about what the truth is better than a potentially less complex meaning does, it eventually is less costly. If we thus assume that alternative scalar meanings are weighed up against each other, the implicature must be computed in any case, that is mandatorily. Note by the way that removing uncertainty may be seen as a discourse factor: answering a question more fully makes for a successful discourse. Thus, Singh's proposal supports our earlier observations about the paramount role of the discourse in the realm of alternatives.

Apart from discourse, there are several other factors that might play a role for the (non-)automatic computation of scalar implicatures, amongst them factors pertaining to the mental or memory capacities of the language users. For example, De Neys and Schaeken (2007) show that participants compute fewer scalar implicatures when they are under a higher processing load (e.g., Bott and Noveck, 2004; Marty et al., 2013), which might be a result of the strengthened meaning not actually being activated. Feng and Cho (2019)[RT] demonstrate that non-native speakers in contrast to native speakers do not compute indirect scalar implicatures (e.g., *not always* → *sometimes*), which might be due to working memory limitations or insufficient linguistic competence in L2. An explanation in terms of linguistic competence would be compatible with findings for children, for whom it has been shown that they do not compute implicatures if they do not know the linguistic expressions denoting the relevant alternatives, which is often the case for quantifiers (e.g., Barner et al., 2011; Horowitz et al., 2018), or if they do not perceive an alternative as relevant (Skordos and Papafragou, 2016). However, Tieu et al. (2019)[RT] observed that neither all children nor all adults compute direct scalar implicatures, so the competence explanation might not be sufficient.

Rees and Bott (2018) show that more implicatures are computed if the alternatives are primed. This is compatible with the assumption that scalar alternatives are not activated obligatorily, but if they are activated—i.e. made salient—the probability of computing an implicature increases. Another way to increase the salience of alternatives is to use prosodic prominence. As mentioned in the *Introduction*, Franke et al. (2017) and other work show that the rate of scalar implicatures increases if the scalar term is prominent. Relatedly, Gotzner (2019) shows that the rate of inference computations (exhaustivity implicatures and additive presuppositions) increases in the presence of a contrastive focus accent. She argues that the accent increases the salience of alternatives and therefore the likelihood of an inference being derived.

Overall, the evidence suggests that alternative activation might not be mandatory although the final answer to this question might depend on the particular type of alternative, i.e. the linguistic domain or subdomain. Overall, factors like prosodic prominence indicating focus/salience, or priming, which also is associated with salience, seem to increase the likelihood that alternatives are activated. Finally, the cost of alternative activation is strongly influenced by contextual factors. In the next subsection, we will focus on those cases where alternatives are activated and look more closely at the time course of this activation.

The Time Course of the Activation of Alternatives During Processing

Assuming that alternatives are activated as part of the language comprehension process, two main questions arise: When do these alternatives become activated and when do they start to influence the unfolding representation of the utterance? Theories addressing these questions give rather different answers depending on the linguistic (sub)domain. For negation and counterfactuals, some accounts assume that the alternative reading may be activated (simulated) before the reading corresponding to the facts. For focus, it is assumed that a focused expression—which itself becomes part of the discourse representation immediately—triggers the activation of, and/or the search for alternatives in the context. Still, findings on the time course are often contradictory, as we will see below.

Regarding negation alternatives, the time course is a matter relevant to two-step models. Kaup and Zwaan (2003) and Kaup et al. (2006), who argued for two-step simulation of the respective situations, such that the simulation of the affirmative situation (p) precedes the simulation of the negative situation ($\neg p$), argue on the basis of their experimental evidence that the tipping point from representing p to representing $\neg p$ must occur about 750 ms after having heard or read a negated statement. Hasson and Glucksberg (2006), who investigated the potential of negated metaphors to prime a word related to the p vs. $\neg p$, put the tipping point at a time between 500 and 1,000 ms, thus supporting Kaup et al.'s assumptions. Hasson and Glucksberg observed that initially, recognition of words related to p is facilitated, and starting from 500 to 1,000 ms, recognition of words related to $\neg p$ is facilitated. Tian et al. (2016), who—recall from the section *Alternatives in Different Linguistic Domains*—present data showing that given an appropriate QuD both representations are activated in parallel, find that the positive situation is available for about 900 ms, in accordance with the time estimates given by Kaup et al. or Hasson and Glucksberg.

The time course of the activation of negation alternatives in counterfactuals was investigated inter alia by Ferguson et al. (2008). They presented participants with the negative antecedent of a counterfactual such as *If cats were not carnivores*, followed by a consequent clause consistent with either the real-world or the counterfactual reality, for instance *families could feed their cat a bowl of carrots* … *Carrots* is the critical word which is consistent with the counterfactual reality (cats are not carnivores). Evidence from eye movements and event-related brain potentials, which were recorded while participants were reading the consequent clause, suggests that the real-world representation was active at the critical word and up to two words further downstream the sentence. Only then did the representation shift. Thus, the counterfactual alternative only becomes available after the real-world alternative has been rejected. Similar findings are reported by de Vega and Urrutia (2012) in a study using event-related brain potentials. These authors claim that the real-world representation is available for about 500 ms, but has faded away after 1,500 ms. See Byrne (2016), Kulakova and Nieuwland (2016a) and Ferguson (2019) for recent reviews of processing aspects of counterfactuals.

Regarding the activation of scalar alternatives, an important insight comes from a trio of studies by different authors, but building on one another: Huang and Snedeker (2009), see also Huang and Snedeker (2011), carried out an eye-tracking study in which participants heard sentences like *Click on the girl who has some of the [ITEMS]* while looking at a visual world display. In the display, there were a girl with some but not all items from the depicted totality next to her, a boy who had the rest of these items, and a girl who had the totality of a set of different items next to her. The critical items next to the girls had an overlapping phonological onset. Thus, eye movements up to and during the first syllable were informative about whether participants entertained the strengthened meaning *some-but-not-all* or the literal meaning *some-and-possibly-all*. The results suggest that the strengthened meaning was available about 800–1,000 ms later than the literal interpretation. In a similar study, Grodner et al. (2010) provided more supporting context for the strengthened meaning, which was then available immediately. Degen and Tanenhaus (2016) carried out two studies using pictures of a gum ball machine (with targets like *You have some of the orange gum balls*). In one of their experiments, they followed the conditions set by Huang and Snedeker (2009) and in the other, those set by Grodner et al. (2010). They replicated both patterns but argue that the sum of the evidence supports immediate availability of the scalar implicature and hence, the co-activation of both alternatives.

Turning to focus, Gotzner and Spalek (2019) in a recent review compare the time course of processing for utterances with a prosodically marked focus vs. utterances where in addition a focus particle associates with the focus. The authors report findings supporting the assumption that mere prosodic focus marking causes an immediate activation of all sorts of related concepts, not just focus alternatives (*contrastive* alternatives in the section *Focus and Contrastive Topics*). However, there also is evidence suggesting immediate activation of focus alternatives only (Braun and Tagliapietra, 2010; Braun and Biezma, 2019[RT]; Yan and Calhoun, 2019[RT]). As time passes, only focus alternatives remain activated. Evidence comes inter alia from Husband and Ferreira (2016), who find delays of 750 ms between the presentation of a constituent prosodically marked for focus and a potential alternative, and from Gotzner et al. (2013), who report a similar effect after a delay of 2,000 ms, but only if the delay has been filled with linguistic material, that is a sentence continuing the narrative, not for silent delays. After a matter of minutes and lasting at least up to a day, alternatives of a prosodically marked focus are recalled better from memory than alternatives which were not prosodically marked (Fraundorf et al., 2010; Fraundorf et al., 2013; Tjuka et al., 2020; Koch and Spalek, 2021).

Utterances containing focus particles do not show early effects, but the particles counteract the online effects observed for focus marked by prosody only, like the facilitated visual recognition of words denoting these alternatives. Gotzner et al. (2016) presented words denoting an alternative about 2 s after a focused element and participants had to decide whether this word had occurred in the sentence (the correct answer was *no*). This decision was made more slowly in sentences with a focus particle than in sentences without a particle. Gotzner and Spalek (2019) assume that the presence of a focus particle triggers an active search for relevant alternatives and that this search causes activated elements to compete, which leads to interference during processing. Eventually though, as in the case for prosodic prominence, focus particles improve memory for alternatives (Spalek et al., 2014).

Summarizing the discussion in the section *The Activation Process*, we found that for some phenomena involving alternatives, the alternatives seem to be available very early, even immediately, whereas for other phenomena, the activation of alternatives is delayed, and there are even findings (for negation and counterfactuals) suggesting that the expressed meaning is available later than the alternative, which needs to be suppressed in processing. Our short review has shown that there is no clear divide by linguistic domain (negation, scalars, focus, etc.). The most striking observation is that for all investigated phenomena, there is at least one study suggesting the immediate availability of an alternative or, in the case of negation and counterfactuals, the immediate availability of the negative situation/the counterfactual world. As Degen and Tanenhaus (2016) demonstrate, the exact experimental details play an important role. Even more important seems to be the role of context. In psycholinguistic experiments, stimuli are often presented shorn of any context to allow for better comparisons between conditions. However, this may render the stimuli highly unnatural. As we saw time and again in the section *Alternatives in Different Linguistic Domains*, context is crucial in the interpretation and relevance of alternatives. If our aim is to understand real-time processing of alternatives, we will have to resort to studies embedding stimuli in naturalistic contexts to closely mimic the way these alternatives are encountered in everyday language use.

Alternatives in the Lab

For experimental studies, it is always challenging to operationalize the main concepts. Oftentimes, a researcher has to make choices that are justified more by experimental design and considerations of doability than by the theory. In this section, we discuss how alternatives can be, and have been operationalized in experiments and what influence this may have on experimental results, focusing on the contributions to this Research Topic.

A first operationalization choice concerns whether alternatives are contextually given or not, and if they are given, whether this is through the linguistic or the situational context, for example by visual co-presence. Doyle et al. (2019)[RT] established alternative sets by placing pairs of (toy) objects on a table and naming them. In a subsequent shopping task involving negative and positive instructions, one object was the target and the other its alternative. Thus, the alternatives were operationalized by co-presence in the situation context. To illustrate, the objects on the table could be an orange and a coconut, and the instruction could be: *The next item is not the orange*, which creates the propositional alternative set *{The next item is not the orange; The next item is the orange}*. Given the context, an inferential step is necessary from *not-the orange* to *the-coconut*.

Both the visual and the linguistic context were manipulated by Kronmüller and Noveck (2019)[RT] in their study of alternatives in relation to conventionalization (section *Negation and Counterfactuals*). The alternatives were determined through co-presence in the current pictorial display and through displays presented previously.

Discourse context as a provider of alternatives is central in Grubic and Wierzba (2019)[RT], who explored the requirements for alternatives for the interpretation of the German focus particle *auch* ('too'). The authors found that alternatives are most likely propositions but that these propositions need not necessarily be salient in the discourse context since comprehenders go to great lengths to identify—and accommodate—relevant propositions. Destruel et al. (2019)[RT] in their study on clefts also provide alternatives in the linguistic context. Overall, this operationalization choice is closest to Rooth's (1992) assumption that focus interpretation introduces an anaphoric variable, which requires an antecedent in the preceding discourse.

Often, visual displays are not used for actually introducing alternatives but for testing what the mental representation of a listener might look like, for instance in Orenes et al. (2019)[RT] and Braun and Biezma (2019)[RT]. The authors presented what they thought the participants' mental representation of alternatives might be in pictorial form or as words on the screen. For the counterfactuals tested by Orenes et al. (*If there had been oranges, there would have been pears*), the actual world is one without oranges and without pears ($\neg p \& \neg q$). This is the representation the participants were expected to form. In the corresponding target picture, oranges and pears were crossed out. The target alternative picture corresponding to the counterfactual world was one with oranges and pears ($p \& q$). As the authors highlight, a distractor picture, which contained apples and strawberries, would also have been consistent with $\neg p \& \neg q$, i.e. *no oranges, no pears*. However, their data suggest that participants preferred an explicit cancellation, as depicted in crossed out oranges and pears. Thus, the mental representation of the alternative set is essentially one of negation alternatives.

In Braun and Biezma's study on contrastively marked topics, the display consisted of different words, one of which was an alternative to the sentence subject. This alternative had been determined empirically in a *not X, but Y* task where participants continued sentence fragments like *Not the gymnast had gotten blisters but the* …. Thus, the authors tried to predict a likely relevant alternative and presented this in the visual environment. If participants look at the alternative more often than at controls, the authors conclude that the alternative has been activated. One caveat in these types of design is that one cannot know whether participants would have activated the particular alternative had it not been presented in the context: the activation could have been triggered retro-actively by the visual presentation. A similar problem arises with lexical decision tasks that are employed to gauge alternative activation through a contrastive accent (here: Yan and Calhoun, 2019[RT]). A useful way to think about focus alternative activation is to assume that focus creates a placeholder for alternatives.[2] This placeholder can be filled either anaphorically from elements in the preceding context, or it can be linked with a likely candidate that is presented after the fact.

Visual presentation of alternatives as a means to find out about participants' mental representation is also employed by Kochari et al. (2020)[RT], who investigated the processing of generic sentences with bare plural subjects. As mentioned in the section *Generics and Plural Definites*, they argue that for the interpretation of generics, three different types of alternatives are important, two of which they test in their experiments: alternatives relevant for the absolute reading of generics (predicate alternatives), and alternatives relevant for the relative reading (subject alternatives). In the experiments, two pictures were presented: one for the target sentence (Beetle type A mostly with dots), and one depicting an alternative subject without the property at issue (Beetle type B without dots). The authors found that the alternative picture indeed was taken into account to judge the truth of the generic, albeit by only part of the participants.

Sometimes, alternative sets are not provided in the experimental setup but are assumed to be created through inferential processes on the basis of characteristics of the expressions for which alternatives are relevant. For instance, if a scalar term like *some* is presented, participants are supposed to infer the strengthened meaning on the basis of a linguistically determined scale. Hence, the alternative set and its members need not be experimentally manipulated. Instead, what is usually manipulated are the combinations of utterances and pictures that are used to assess whether participants have interpreted an implicature or not. An interesting facet (e.g., Feng and Cho, 2019[RT]) are negated scalar expressions due to the combination of scalar and negation alternatives. The alternative set in this case seems to consist of the non-negated (affirmative) semantic meaning, as well as the negative semantic meaning and its strengthened meaning, in line with two-step theories of negation processing.

Experimental displays without 'explicit' alternatives have also been used for the investigation of homogeneity. Tieu et al. (2019)[RT] presented pictures of a set of the same objects (e.g., hearts) either in a single color or in different colors. Participants judged the felicity of sentences with plural definites like *The hearts are yellow* for a picture of red and yellow hearts. The experiment was carried out in French with French-speaking children. Crucially, the authors employed a set of critical comparisons to determine how the plural definites are interpreted. These included sentences with the scalar expressions *none*, *some*, and *all*. Thus, scalar alternatives do not become relevant through direct, explicit juxtaposition. Still, they are obviously contextually present in this setup. Similarly, Beltrán et al. (2019)[RT] and Haase et al. (2019)[RT] in their investigations of neural processing mechanisms during negation processing do not use explicit alternatives. However, Haase et al. used stimuli which had alternatives across experimental trials as they contained co-hyponyms to the hyperonym *professions*. So this study, too, involved a discourse context providing alternatives.

The studies by Haase et al. (2019)[RT] and Yan and Calhoun (2019)[RT] illustrate another design choice. As we saw, researchers often provide possible co-hyponyms in order to probe whether alternatives are active. There is nothing in the definition of alternatives that requires this relationship and, in fact, for focus alternatives, a number of studies have tested explicitly whether alternatives have to be co-hyponyms (Gotzner, 2015; Kim et al., 2015; Jördens et al., 2020): The answer is *no*. Still, it is interesting to ask why co-hyponymy is often used as a convenient shortcut in the operationalization of alternatives. First, Rooth's focus semantic value is often 'translated' for empirical purposes as the set of propositions obtained by replacing the focused element with an alternative of the same semantic type such that the proposition is still sensible. Co-hyponyms are well suited for this: If I can carry out an action (cut, squeeze, bake) with an individual, I can usually carry out that action with individuals which are denoted by co-hyponyms. Co-hyponyms are used particularly often for testing contrastive focus. Repp (2010)

[2]This idea goes back to Steven Crain.

discusses the notion of contrast and cites Kiss's (1998) requirement that contrastive focus needs a complementary alternative set with clearly identifiable elements. This means that the alternative set has to be closed and that alternatives need to be mentioned in the context. Co-hyponymy might contribute to identifiability: If a given hyperonym has only very few hyponyms, the alternative set is easily identifiable. The closed-set argument might apply to all co-hyponym relationships, but it may be easier for some than for others.

Alternatives for All? Individual Differences

An important aim in any field of research is to formulate generalizable conclusions that hold for a well-defined population. However, it has become clear that "all language users" or even "all adult native language users" defines the population too broadly. While it is relatively uncontroversial that native language speakers differ from language learners and children differ from adults, the insight has gradually emerged that even within the group of native, adult language users, subgroups can be found who process a given linguistic phenomenon differently (Kidd et al., 2018). Thus, the challenge becomes to describe and understand the dimensions along which the population is grouped. For instance, in the field of scalar implicatures, so-called logical comprehenders and pragmatic responders emerge. While logical responders do not draw the implicature, interpreting, for example, an utterance containing the scalar term *some* as *some-and-possibly-all*, pragmatic responders do, interpreting the term as meaning *some-and-not-all* (Noveck and Posada, 2003; Bott and Noveck, 2004 (Exp. 3); Bott et al., 2012; Tomlinson and Bott, 2013; Spychalska et al., 2016). In this final section, we investigate if groups can be identified also for the processing of other alternative-related phenomena.

Five of the contributions in this Research Topic address individual differences, either in passing or as a research question in its own right. They fall into two categories with regard to how they look at individual difference. Either groups were defined beforehand (children vs. adults: Doyle et al.; L1 vs. L2 speakers: Feng and Cho) or emerging groups were described (Orenes et al., Kochari et al.). Tieu et al. are a special case in that they hypothesized the existence of three groups, but did not know how exactly these might be represented in the population they tested.

Starting with the first category, recall that Doyle et al. (2019)[RT] investigated negation processing with a shopping task. Adults and children selected one item from a set of two and put it in a shopping cart. Both their response latencies and their eye movements were measured. The authors observed that both adults and children looked more often at the non-target when hearing a negative than a positive sentence (e.g., for *The next item is not an apple* they looked more often at the apple than they looked at the non-apple for *The next item is an apple*.). However, children were slower in their responses. The authors conclude that children's processing of negation is not yet as effective as that of adults (cf. Nordmeyer and Frank, 2014). Feng and Cho (2019)[RT] compared direct and indirect scalar implicatures (*sometimes* → not always; *not always* → sometimes) for native speakers of English and L2 English learners with a covered box paradigm. Participants were presented with a visible picture and an invisible one (the covered box) and chose either, depending on the meaning they assign to a sentence they hear. The groups behaved remarkably similar. The only significant difference was obtained when a no-inference picture was chosen for the indirect scalar implicature. In this case, non-native speakers were more likely to select the visible picture (i.e., to suspend the inference) than native speakers. Both these acquisition studies support the assumption that there exists a developmental path to a certain manner of processing. Children and L2 learners differed from adults/native speakers in the most effortful condition only, suggesting that they had not yet reached mastery with the computation of these meanings.

As mentioned, Tieu et al. (2019)[RT] predefine three groups in their study on plural definites but what they find is on the one hand different groups, and on the other hand unexpected individual differences. Recall that adults display a truth-value gap for the use of plural definites in non-homogeneous situations (*The beetles are red* is neither true nor false if only some beetles are red; section *Generics and Plural Definites*; Križ and Chemla, 2015). Tieu et al. reason that young children might fall into the following groups: The "homogeneity group," whose performance equals that of adults, the "existential group," who accepts the affirmative description but rejects the negative one (→ there are some beetles that are red), and the "universal group," who accepts the negative description but rejects the affirmative one (→ it is not the case that all the beetles are red). Tieu et al. observe that adults are not as uniform a group as was previously assumed: a small number of adults interpreted the utterances universally. Children fell either into the "homogeneity" or the "existential group." With even more fine-grained group assignments, the authors identified three groups of children: those who interpreted the plural existentially and did not compute scalar implicatures, those who made the homogeneity assumption and computed scalar implicatures (= adult-like), and those who have adult-like homogeneity readings while not computing scalar implicatures. An interesting question is whether these groups in the child population will all develop into the adult "homogeneity group" or whether a certain group is more likely to end up interpreting these utterances universally, just as a small subgroup of adults did.

In the study by Orenes et al. (2019)[RT] on counterfactuals, the focus on individual differences also was *post-hoc* and it was driven by the observation that confidence intervals by participants were much larger than those by items. As mentioned in the section *The Time Course of the Activation of Alternatives During Processing*, participants' looks to the picture representing the factual world (*¬p&¬q*) and to the picture representing the counterfactual world (*p&q*) started to rise quickly upon presentation of a counterfactual. After indicative conditionals, in contrast, only looks to the *p&q* picture were observed. A post-hoc analysis revealed that one group of participants showed exactly the same looking behavior for indicative conditionals and counterfactuals: they looked more at the *p&q* picture. The other group looked at both pictures (or only at *¬p&¬q*). Orenes et al. argue that the participants who only looked at the *p&q* picture for both types of conditionals did not retrieve the correct meaning for the counterfactual target sentence. Kulakova and Nieuwland (2016b) have traced some of the individual differences in processing counterfactuals back to differences in the participants' abilities to understand the communicative intentions of others.

Finally, Kochari et al. (2020)[RT] observed considerable variation between individual participants in their study on generics. As briefly mentioned above, they found that the alternative picture required for the relative meaning of the generic (= subject alternatives) was only considered by one group of participants. The other group interpreted the generic with an absolute meaning, independently of the presence of a picture which would license the relative reading.

In sum, there are several independent challenges when investigating individual variation. First, individuals who do not understand a target structure as intended need to be eliminated from the sample and, ideally, an explanation needs to be found for why they do not process the structure as intended. Second, researchers need to determine whether the remaining individuals all reach their interpretation in the same way or whether there are different pathways to (correct) understanding. And there are many more potential challenges surrounding individual differences in language processing, whose scope we are only beginning to understand.

CONCLUSION

This review has discussed the notion of alternatives in meaning interpretation in several linguistic domains. We have argued that there is good reason to believe that there are indeed different "types" of alternatives, which are subject to different conditions, and which may interact. Although discourse context is crucial for all types of alternatives, it is probably only semantically required for focus alternatives, as focus alternatives are licensed via anaphoric context conditions. However, we saw that the accommodation of focus alternatives is possible and that in the absence of context, lexical-semantic relations like co-hyponymy are exploited. Furthermore, the desire to create coherent discourses also guides the search and selection of alternatives. We have also argued that the precise mechanisms of the on-line activation of alternatives is highly controversial and that the evidence often is inconclusive. Factors influencing this state of affairs are on the one hand the operationalization of the notion of alternatives in experimental settings, and on the other hand individual variation, which poses challenges both for descriptive generalizations and for theoretical modeling. We hope to have made a contribution to the field of alternatives research that inspires future work which intersects aspects of grammar and processing to an even greater extent than the last decade has already seen.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement GAP-677742, awarded to KS.

REFERENCES

Alonso-Ovalle, L. (2009). Counterfactuals, Correlatives, and Disjunction. *Linguist. Philos.* 32 (2), 207–244. doi:10.1007/s10988-009-9059-0

Anderson, S., Huette, S., Matlock, T., and Spivey, M. (2010). "On the Temporal Dynamics of Negated Perceptual Simulations," in *Meaning, Form, and Body*. Editors F. Parrill, V. Tobin, and M. Turner (Stanford, CA: Stanford University Press), 1–20.

Arregui, A. (2020). "Counterfactuals: 'If Kangaroos Had No Tails . . .'," in *The Wiley Blackwell Companion to Semantics*. Editors D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. Zimmermann (Hoboken, NJ: Wiley), 1–33.

Arroyo, F. V. (1982). Negatives in Context. *J. Verb. Learning Verb. Behav.* 21 (1), 118–126. doi:10.1016/S0022-5371(82)90500-X

Asher, N., and Lascarides, A. (1998). The Semantics and Pragmatics of Presupposition. *J. Semant.* 15 (3), 239–300. doi:10.1093/jos/15.3.239

Asher, N., and Morreau, M. (1995). "What Some Generic Sentences Mean," in *The Generic Book*. Editors G. Carlson and F. J. Pelletier (Chicago, IL: Chicago University Press), 300–339.

Atlas, J., and Levinson, S. (1981). "It-Clefts, Informativeness, and Logical Form: Radical Pragmatics (Revised Standard Version)," in *Radical Pragmatics*. Editor P. Cole (New York, NY: Academic Press), 1–61.

Autry, K. S., and Levine, W. H. (2012). Activation of negated and non-negated entities. *J. Pragmatics* 44, 1474–1485.

Bar-Lev, M. (2018). *Free Choice, Homogeneity, and Innocent Inclusion*. [dissertation]. Jerusalem (Israel): Hebrew University of Jerusalem.

Barner, D., Brooks, N., and Bale, A. (2011). Accessing the Unsaid: The Role of Scalar Alternatives in Children's Pragmatic Inference. *Cognition* 118 (1), 84–93. doi:10.1016/j.cognition.2010.10.010

Beaver, D., and Clark, B. (2008). *Sense and Sensitivity: How Focus Determines Meaning*. Cambridge, United Kingdom: Wiley-Blackwell.

Beltrán, D., Morera, Y., García-Marco, E., and Vega, M. d. (2019). Brain Inhibitory Mechanisms are Involved in the Processing of Sentential Negation, Regardless of its Content. Evidence From EEG Theta and Beta Rhythms. *Front. Psychol.* 10, 1782. doi:10.3389/fpsyg.2019.01782

Benz, A., and Gotzner, N. (2020). Embedded Implicature: What Can be Left Unsaid?. *Linguist. Philos.* 1–32. doi:10.1007/s10988-020-09310-x

Benz, A., and Stevens, J. (2018). Game-Theoretic Approaches to Pragmatics. *Annu. Rev. Linguist.* 4, 173–191. doi:10.1146/annurev-linguistics-011817-045641

Benz, A., and Van Rooij, R. (2007). Optimal Assertions, and What They Implicate. A Uniform Game Theoretic Approach. *Topoi* 26 (1), 63–78. doi:10.1007/s11245-006-9007-3

Bott, L., Bailey, T. M., and Grodner, D. (2012). Distinguishing Speed From Accuracy in Scalar Implicatures. *J. Mem. Lang.* 66 (1), 123–142. doi:10.1016/j.jml.2011.09.005

Bott, L., and Noveck, I. A. (2004). Some Utterances are Underinformative: The Onset and Time Course of Scalar Inferences. *J. Mem. Lang.* 51 (3), 437–457. doi:10.1016/j.jml.2004.05.006

Braun, B., and Biezma, M. (2019). Prenuclear L*+H Activates Alternatives for the Accented Word. *Front. Psychol.* 10, 1993. doi:10.3389/fpsyg.2019.01993

Braun, B., and Tagliapietra, L. (2010). The Role of Contrastive Intonation Contours in the Retrieval of Contextual Alternatives. *Lang. Cogn. Process.* 25 (7–9), 1024–1043. doi:10.1080/01690960903036836

Braun, B., Asano, Y., and Dehé, N. (2018). When (not) to look for contrastive alternatives: the role of pitch accent type and additive particles. *Lang. Speech* 5, 1474–1485. doi:10.1177/0023830918814279

Breheny, R., Katsos, N., and Williams, J. (2006). Are Generalised Scalar Implicatures Generated by Default? An On-Line Investigation Into the Role of Context in Generating Pragmatic Inferences. *Cognition* 100 (3), 434–463. doi:10.1016/j.cognition.2005.07.003

Breheny, R. (2019). "Scalar Implicatures," in *The Oxford Handbook of Experimental Semantics and Pragmatics*. Editors C. Cummins and N. Katsos (Oxford, United Kingdom: Oxford University Press), 38–61.

Büring, D. (2016a). "(Contrastive) Topic," in *The Oxford Handbook of Information Structure*. Editors C. Féry and S. Ishihara (Oxford, United Kingdom: Oxford University Press), 64–85.

Büring, D. (2016b). *Intonation and Meaning*. Oxford, United Kingdom: Oxford University Press.

Büring, D., and Križ, M. (2013). It's That, and That's it! Exhaustivity and Homogeneity Presuppositions in Clefts (and Definites). *Semantics and Pragmatics* 6, 1–29. doi:10.3765/sp.6.6

Büring, D. (2003). On D-Trees, Beans, and B-Accents. *Linguist. Philos.* 26 (5), 511–545. doi:10.1023/A:1025887707652

Byrne, R. M. J. (2016). Counterfactual Thought. *Annu. Rev. Psychol.* 67, 135–157. doi:10.1146/annurev-psych-122414-033249

Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.

Carlson, G. (1977). Reference to Kinds in English. [dissertation]. Amherst (MA): University of Massachusetts.

Carpenter, P. A., and Just, M. A. (1975). Sentence Comprehension: A Psycholinguistic Processing Model of Verification. *Psychol. Rev.* 82 (1), 45–73. doi:10.1037/h0076248

Chemla, E., and Bott, L. (2014). Processing Inferences at the Semantics/Pragmatics Frontier: Disjunctions and Free Choice. *Cognition* 130 (3), 380–396. doi:10.1016/j.cognition.2013.11.013

Chemla, E., Cummins, C., and Singh, R. (2016). Training and Timing Local Scalar Enrichments Under Global Pragmatic Pressures. *J. Semant.* 34, 107–126. doi:10.1093/jos/ffw006

Chemla, E., and Singh, R. (2014). Remarks on the Experimental Turn in the Study of Scalar Implicature, Part I. *Lang. Linguist. Compass* 8, 373–386. doi:10.1111/lnc3.12081

Chemla, E. (2009). Universal Implicatures and Free Choice Effects: Experimental Data. *Semantics and Pragmatics* 2, 1–33. doi:10.3765/sp.2.2

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., and Sperber, D. (2008). Making Disjunctions Exclusive. *Q. J. Exp. Psychol.* 61 (11), 1741–1760. doi:10.1080/17470210701712960

Chierchia, G. (2006). Broaden Your Views: Implicatures of Domain Widening and the "Logicality" of Language. *Linguist. Inq.* 37 (4), 535–590. doi:10.1162/ling.2006.37.4.535

Chierchia, G. (2011). "Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface," in *Structures and beyond*. Editor A. Belletti (Oxford: Oxford University Press), 39–103.

Chierchia, G., Fox, D., and Spector, B. (2012). "The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics," in *Semantics: An International Handbook of Natural Language Meaning 3*. Editors C. Maienborn, K. von Heusinger, and P. Portner (Berlin, Germany: De Gruyter), 2297–2232.

Ciardelli, I., Groenendijk, J., and Roelofsen, F. (2018). *Inquisitive Semantics*. Oxford, United Kingdom: Oxford University Press.

Clark, H. H., and Chase, W. G. (1972). On the Process of Comparing Sentences Against Pictures. *Cogn. Psychol.* 3 (3), 472–517. doi:10.1016/0010-0285(72)90019-9

Claus, B., Frühauf, F., and Krifka, M. (2019). "Interpreting Polarity-Ambiguous Propositional Anaphors with Negative Antecedents: Some Experimental Results," in Proceedings of Sinn und Bedeutung 23:1. Editors M. T. Espinal, E. Castroviejo, M. Leonetti, L. McNally, and C. Real-Puigdollers (Cerdanyola del Vallès, Spain: Autonomous University of Barcelona), 267–283.

Cohen, A. (2004). Generics and Mental Representations. *Linguist. Philos.* 27 (5), 529–556. doi:10.1023/B:LING.0000033851.25870.3e

Cohen, A. (1999). *Think Generic! The Meaning and Use of Generic Sentences*. Stanford, CA: CSLI Publications.

Constant, N. (2014). Contrastive Topic: Meanings and Realizations. [dissertation]. Amherst (MA): University of Massachusetts.

Corblin, F. (1991). *Presupposition and Discourse Context*. Paris, France: CNRS.

Cornish, E. R. (1971). Pragmatic Aspects of Negation in Sentence Evaluation and Completion Tasks. *Br. J. Psychol.* 62 (4), 505–511. doi:10.1111/j.2044-8295.1971.tb02065.x

Cremers, A., and Chemla, E. (2014). "Direct and Indirect Scalar Implicatures Share the Same Processing Signature," in *Pragmatics, Semantics and the Case of Scalar Implicatures*. Editor S. P. Reda (London, United Kingdom: Palgrave Macmillan), 201–227.

Crnič, L., Chemla, E., and Fox, D. (2015). Scalar Implicatures of Embedded Disjunctions. *Nat. Lang. Semant.* 23, 271–305. doi:10.1007/s11050-015-9116-x

Dahan, D., Tanenhaus, M. K., and Chambers, C. G. (2002). Accent and Reference Resolution in Spoken-Language Comprehension. *J. Mem. Lang.* 47 (2), 292–314. doi:10.1007/s10936-014-9306-610.1016/s0749-596x(02)00001-3

Dale, R., and Duran, N. D. (2011). The Cognitive Dynamics of Negated Sentence Verification. *Cogn. Sci.* 35 (5), 983–996. doi:10.1111/j.1551-6709.2010.01164.x

De Neys, W., and Schaeken, W. (2007). When People are More Logical under Cognitive Load. *Exp. Psychol.* 54 (2), 128–133. doi:10.1027/1618-3169.54.2.128

De Veaugh-Geiss, J. P., Tönnis, S., Onea, E., and Zimmermann, M. (2018). That's Not Quite it: An Experimental Investigation of (Non)exhaustivity in Clefts. *Semantics and Pragmatics* 11, 3. doi:10.3765/sp.11.3

De Vega, M., and Urrutia, M. (2012). Discourse Updating After Reading a Counterfactual Event. *Psicológica* 33 (2), 157–173.

Degen, J., and Tanenhaus, M. K. (2016). Availability of Alternatives and the Processing of Scalar Implicatures: A Visual World Eye-Tracking Study. *Cogn. Sci.* 40 (1), 172–201. doi:10.1111/cogs.12227

Destruel, E., Beaver, D. I., and Coppock, E. (2019). It's not what you expected! The surprising nature of cleft alternatives in French and English. *Front. Psychol.* 10, 1400. doi:10.3389/fpsyg.2019.01400

Dowty, D. (1987). "Collective Predicates, Distributive Predicates and All," in Proceedings of the Third Eastern States Conference on Linguistics (ESCOL). Editors A. Miller and Z.-S. Zhang (Columbus, OH: Ohio State University), 97–115.

Doyle, A. W., Friesen, K., Reimer, S., and Pexman, P. M. (2019). Grasping the Alternative: Reaching and Eyegaze Reveal Children's Processing of Negation. *Front. Psychol.* 10, 1227. doi:10.3389/fpsyg.2019.01227

Dudschig, C., Maienborn, C., and Kaup, B. (2016). Is There a Difference Between Stripy Journeys and Stripy Ladybirds? The N400 Response to Semantic and World-Knowledge Violations During Sentence Processing. *Brain Cogn.* 103, 38–49. doi:10.1016/j.bandc.2016.01.001

Feng, S., and Cho, J. (2019). Asymmetries Between Direct and Indirect Scalar Implicatures in Second Language Acquisition. *Front. Psychol.* 10, 877. doi:10.3389/fpsyg.2019.00877

Ferguson, H. (2019). "Counterfactuals," in *The Oxford Handbook of Experimental Semantics and Pragmatics*. Editors C. Cummins and N. Katsos (Oxford, United Kingdom: Oxford University Press), 123–142.

Ferguson, H. J., Sanford, A. J., and Leuthold, H. (2008). Eye-Movements and ERPs Reveal the Time Course of Processing Negation and Remitting Counterfactual Worlds. *Brain Res.* 1236, 113–125. doi:10.1016/j.brainres.2008.07.099

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., and Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology* 20, 400–409.

Fodor, J. D. (1970). The Linguistic Description of Opaque Contexts. [dissertation]. Cambridge (MA): Massachusetts Institute of Technology.

Fox, D., and Katzir, R. (2011). On the Characterization of Alternatives. *Nat. Lang. Semant.* 19 (1), 87–107. doi:10.1007/s11050-010-9065-3

Fox, D., and Spector, B. (2018). Economy and Embedded Exhaustification. *Nat. Lang. Semant.* 26 (1), 1–50. doi:10.1007/s11050-017-9139-6

Frank, M. C., and Goodman, N. D. (2014). Inferring Word Meanings by Assuming that Speakers are Informative. *Cogn. Psychol.* 75, 80–96. doi:10.1016/j.cogpsych.2014.08.002

Frank, M. C., and Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science* 336 (6084), 998. doi:10.1126/science.1218633

Franke, M. (2011). Quantity Implicatures, Exhaustive Interpretation, and Rational Conversation. *Semantics and Pragmatics* 4 (1), 1–82. doi:10.3765/sp.4.1

Franke, M., Schlotterbeck, F., and Augurzky, P. (2017). Embedded Scalars, Preferred Readings and Prosody: An Experimental Revisit. *J. Semant.* 34 (1), 153–199. doi:10.1093/jos/ffw007

Fraundorf, S. H., Benjamin, A. S., and Watson, D. G. (2013). What Happened (and What Did Not): Discourse Constraints on Encoding of Plausible Alternatives. *J. Mem. Lang.* 69 (3), 196–227. doi:10.1016/j.jml.2013.06.003

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (2010). Recognition Memory Reveals Just How CONTRASTIVE Contrastive Accenting Really is. *J. Mem. Lang.* 63 (3), 367–386. doi:10.1016/j.jml.2010.06.004

Fretheim, T. (1992). The Effect of Intonation on a Type of Scalar Implicature. *J. Pragmat.* 18 (1), 1–30. doi:10.1016/0378-2166(92)90105-K

Fălăuş, A. (2020). "Negation and Alternatives: Interaction with Focus Constituents," in *The Oxford Handbook of Negation*. Editors V. Déprez and M. T. Espinal (Oxford, United Kingdom: Oxford University Press), 38–49.

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York, NY: Academic Press.

Geurts, B. (2010). *Quantity Implicatures*. Cambridge, United Kingdom: Cambridge University Press.

Geurts, B., and van der Sandt, R. (2004). Interpreting Focus. *Theor. Linguist.* 30, 1–44. doi:10.1515/thli.2004.005

Geurts, B. (2009). Scalar Implicature and Local Pragmatics. *Mind Lang.* 24 (1), 51–79. doi:10.1111/j.1468-0017.2008.01353.x

Givón, T. (1978). "Negation in Language: Pragmatics, Function, Ontology," in *Pragmatics*. Editor P. Cole (New York, NY: Academic Press), 69–112.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., and Johnson-Glenberg, M. C. (1999). Not Propositions. *Cogn. Syst. Res.* 1 (1), 19–33. doi:10.1016/S1389-0417(99)00004-2

Goodman, N. (1955). *Fact, Fiction and Forecast*. Indianapolis, IN: The Bobbs-Merrill Company.

Gotzner, N. (2017). *Alternative Sets in Language Processing: How Focus Alternatives are Represented in the Mind*. London, United Kingdom: Palgrave Macmillan.

Gotzner, N., and Spalek, K. (2019). The Life and Times of Focus Alternatives: Tracing the Activation of Alternatives to a Focused Constituent in Language Comprehension. *Lang. Linguist. Compass* 13, e12310. doi:10.1111/lnc3.12310

Gotzner, N., Spalek, K., and Wartenburger, I. (2013). "How Pitch Accents and Focus Particles Affect the Recognition of Contextual Alternatives," in Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci). Editors M. Knauff, M. Paun, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Scienece Society), 2434–2439.

Gotzner, N. (2019). The Role of Focus Intonation in Implicature Computation: A Comparison With Only and Also. *Nat. Lang. Semant.* 27 (3), 189–226. doi:10.1007/s11050-019-09154-7

Gotzner, N., Wartenburger, I., and Spalek, K. (2016). The Impact of Focus Particles on the Recognition and Rejection of Contrastive Alternatives. *Lang. Cogn.* 8 (1), 59–95. doi:10.1017/langcog.2015.25

Gotzner, N. (2015). "What's Included in the Set of Alternatives? Psycholinguistic Evidence for a Permissive View," in Proceedings of Sinn und Bedeutung 19. Editors E. Csipak and H. Zeijlstra (Göttingen: University of Göttingen), 252–267.

Gough, P. B. (1965). Grammatical Transformations and Speed of Understanding. *J. Verb. Learning. Verb. Behav.* 4 (2), 107–111. doi:10.1016/S0022-5371(65)80093-7

Grice, H. P. (1975). "Logic and Conversation," in *Syntax and Semantics 3: Speech Acts*. Editors P. Cole and J. L. Morgan (New York, NY: Academic Press), 41–58.

Grice, H. P. (1967). *Logic and Conversation (William James Lectures)*. Cambridge, MA: Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., and Tanenhaus, M. K. (2010). "Some," and Possibly All, Scalar Inferences are Not Delayed: Evidence for Immediate Pragmatic Enrichment. *Cognition* 116 (1), 42–55. doi:10.1016/j.cognition.2010.03.014

Grubic, M., and Wierzba, M. (2019). Presupposition Accommodation of the German Additive Particle Auch (= "Too"). *Front. Commun.* 4, 15. doi:10.3389/fcomm.2019.00015

Haase, V., Spychalska, M., and Werning, M. (2019). Investigating the Comprehension of Negated Sentences Employing World Knowledge: An Event-Related Potential Study. *Front. Psychol.* 10, 2184. doi:10.3389/fpsyg.2019.02184

Hajičová, E. (1996). "Topic-Focus Articulation – A Matter of Langue or Parole? The Case of Negation," in *Theoretical Linguistics and Grammatical Description 138: Papers in Honour of Hans-Heinrich Lieb on the Occasion of His 60th Birthday*. Editor R. Sackmann (Amsterdam, Netherlands: John Benjamins), 167–175.

Halvorsen, P.-K. (1978). The Syntax and Semantics of Cleft Constructions. [dissertation]. Austin (TX): University of Texas at Austin.

Hasson, U., and Glucksberg, S. (2006). Does Understanding Negation Entail Affirmation?. *J. Pragmat.* 38 (7), 1015–1032. doi:10.1016/j.pragma.2005.12.005

Heim, I. (1992). Presupposition Projection and the Semantics of Attitude Verbs. *J. Semant.* 9 (3), 183–221. doi:10.1093/jos/9.3.183

Herburger, E. (2000). *What Counts: Focus and Quantification*. Cambridge, MA: MIT Press.

Hörmann, H. (1971). Semantic Factors in Negation. *Psychol. Forsch.* 35, 1–16. doi:10.1007/BF00424470

Horn, L. (1972). On the Semantic Properties of Logical Operators in English. [dissertation]. Los Angeles (CA): University of California, Los Angeles.

Horn, L. R. (1981). "Exhaustiveness and the Semantics of Clefts," in Proceedings of the Eleventh Annual Meeting of the North Eastern Linguistic Society (NELS). Editors V. E. Burke and J. Pustejovsky (Amherst, MA: University of Massachusetts), 125–142.

Horn, L. R. (2006). "The Border Wars: A Neo-Gricean Perspective," in *Where Semantics Meets Pragmatics*. Editors K. von Heusinger and K. Turner (Amsterdam, Netherlands: Elsevier), 21–48.

Horowitz, A. C., Schneider, R. M., and Frank, M. C. (2018). The Trouble With Quantifiers: Exploring Children's Deficits in Scalar Implicature. *Child. Dev.* 89 (6), e572–e593. doi:10.1111/cdev.13014

Huang, Y. T., and Snedeker, J. (2011). Logic and Conversation Revisited: Evidence for a Division between Semantic and Pragmatic Content in Real-Time Language Comprehension. *Lang. Cogn. Process.* 26 (8), 1161–1172. doi:10.1080/01690965.2010.508641

Huang, Y. T., and Snedeker, J. (2009). Online Interpretation of Scalar Quantifiers: Insight Into the Semantics-Pragmatics Interface. *Cogn. Psychol.* 58 (3), 376–415. doi:10.1016/j.cogpsych.2008.09.001

Hurford, J. (1974). Exclusive or Inclusive Disjunction. *Found. Lang.* 11 (3), 409–411.

Husband, E. M., and Ferreira, F. (2016). The Role of Selection in the Comprehension of Focus Alternatives. *Lang. Cogn. Neurosci.* 31 (2), 217–235. doi:10.1080/23273798.2015.1083113

Ito, K., and Speer, S. R. (2008). Anticipatory Effects of Intonation: Eye Movements During Instructed Visual Search. *J. Mem. Lang.* 58 (2), 541–573. doi:10.1016/j.jml.2007.06.013

Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jacobs, J. (1991). "Negation," in *Semantics: An International Handbook of Contemporary Research*. Editors A. von Stechow and D. Wunderlich (Berlin, Germany: De Gruyter Mouton), 560–596.

Johnson-Laird, P. N., and Tridgell, J. M. (1972). When Negation is Easier Than Affirmation. *Q. J. Exp. Psychol.* 24 (1), 87–91. doi:10.1080/14640747208400271

Jördens, K. A., Gotzner, N., and Spalek, K. (2020). The Role of Non-Categorical Relations in Establishing Focus Alternative Sets. *Lang. Cogn.* 12 (4), 729–754. doi:10.1017/langcog.2020.21

Katzir, R. (2007). Structurally-Defined Alternatives. *Linguist. Philos.* 30 (6), 669–690. doi:10.1007/s10988-008-9029-y

Kaup, B., and Dudschig, C. (2020). "Understanding Negation," in *The Oxford Handbook of Negation*. Editors V. Déprez and M. T. Espinal (Oxford, United Kingdom: Oxford University Press), 634–655. doi:10.1093/oxfordhb/9780198830528.013.33

Kaup, B., Lüdtke, J., and Zwaan, R. A. (2006). Processing Negated Sentences With Contradictory Predicates: Is a Door That is Not Open Mentally Closed?. *J. Pragmat.* 38 (7), 1033–1050. doi:10.1016/j.pragma.2005.09.012

Kaup, B., and Zwaan, R. A. (2003). Effects of Negation and Situational Presence on the Accessibility of Text Information. *J. Exp. Psychol. Learn. Mem. Cogn.* 29 (3), 439–446. doi:10.1037/0278-7393.29.3.439

Kaup, B., Zwaan, R. A., and Lüdtke, J. (2007). "The Experiential View of Language Comprehension: How Is Negation Represented?" in *Higher Level Language Processes in the Brain: Inference and Comprehension Processes*. Editors C. Schmalhofer and C. A. Perfetti (Mahwah, NJ and London, United Kingdom: Lawrence Erlbaum Associates, 255–288.

Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends Cogn. Sci.* 22 (2), 154–169. doi:10.1016/j.tics.2017.11.006

Kim, C. S., Gunlogson, C., Tanenhaus, M. K., and Runner, J. T. (2015). Context-driven Expectations about Focus Alternatives. *Cognition* 139, 28–49. doi:10.1016/j.cognition.2015.02.009

Kiss, K. É. (1998). Identificational Focus Versus Information Focus. *Language* 74 (2), 245–273. doi:10.1353/lan.1998.0211

Koch, X., and Spalek, K. (2021). Contrastive intonation effects on word recall for information-structural alternatives across the sexes. *Mem. Cogn.* doi:10.3758/s13421-021-01174-1

Kochari, A., Van Rooij, R., and Schulz, K. (2020). Generics and Alternatives. *Front. Psychol.* 11, 1274. doi:10.3389/fpsyg.2020.01274

Kounios, J., and Holcomb, P. J. (1992). Structure and Process in Semantic Memory: Evidence From Event-Related Brain Potentials and Reaction Times. *J. Exp. Psychol. Gen.* 121 (4), 459–479. doi:10.1037/0096-3445.121.4.459

Krifka, M. (2008). Basic Notions of Information Structure. *Acta Linguist. Hung.* 55 (3–4), 243–276. doi:10.1556/aling.55.2008.3-4.2

Krifka, M. (1996). Parametrized Sum Individuals for Plural Anaphora. *Linguist. Philos.* 19, 6555–6598. doi:10.1007/bf00632708

Krifka, M. (2013). "Response Particles as Propositional Anaphors," in Proceedings Semantics and Linguistic Theory. Editor Snider, T. (Ithaca, NY: CLC), 1–18.

Križ, M. (2015). Aspects of Homogeneity in the Semantics of Natural Language. [dissertation]. Vienna (Austria): University of Vienna.

Križ, M., and Chemla, E. (2015). Two Methods to Find Truth-Value Gaps and Their Application to the Projection Problem of Homogeneity. *Nat. Lang. Semant.* 23 (3), 205–248. doi:10.1007/s11050-015-9114-z

Križ, M., and Spector, B. (2020). Interpreting Plural Predication: Homogeneity and Non-Maximality. *Linguist. Philos.* 1–48. doi:10.1007/s10988-020-09311-w

Kroll, J. F., and Corrigan, A. (1981). Strategies in Sentence-Picture Verification: The Effect of an Unexpected Picture. *J. Verb. Learning Verb. Behav.* 20 (5), 515–531. doi:10.1016/S0022-5371(81)90147-X

Kronmüler, E., and Noveck, I. (2019). How Do Addressees Exploit Conventionalizations? From a Negative Reference to an *Ad Hoc* Implicature. *Front. Psychol.* 10, 1461. doi:10.3389/fpsyg.2019.01461

Kronmüller, E., and Barr, D. J. (2015). Referential Precedents in Spoken Language Comprehension: A Review and Meta-Analysis. *J. Mem. Lang.* 83, 1–19. doi:10.1016/j.jml.2015.03.008

Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., and Barr, D. (2017). The Positive Side of a Negative Reference: The Delay Between Linguistic Processing and Common Ground. *R. Soc. Open Sci.* 4, 160827. doi:10.1098/rsos.160827

Kulakova, E., and Nieuwland, M. S. (2016b). Pragmatic Skills Predict Online Counterfactual Comprehension: Evidence From the N400. *Cogn. Affect. Behav. Neurosci.* 16, 814–824. doi:10.3758/s13415-016-0433-4

Kulakova, E., and Nieuwland, M. S. (2016a). Understanding Counterfactuality: A Review of Experimental Evidence for the Dual Meaning of Counterfactuals. *Lang. Linguist. Compass* 10 (2), 49–65. doi:10.1111/lnc3.12175

Löbner, S. (1987). "The Conceptual Nature of Natural Language Quantification," in Proceedings of the 87th Debrecen Symposium on Logic and Language. Editors I. Rusza and A. Szabolcsi (Budapest, Hungary: Akadémiai Kiadó), 81–94.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Lewis, D. (1973). *Counterfactuals*. Oxford, United Kingdom: Blackwell.

Löbner, S. (2000). Polarity in Natural Language: Predication, Quantification and Negation in Particular and Characterizing Sentences. *Linguist. Philos.* 23 (3), 213–308. doi:10.1023/A:1005571202592

Lüedtke, J., Friedrich, C. K., De Filippi, M., and Kaup, B. (2008). Event-Related Potential Correlates of Negation in a Sentence-Picture Verification Paradigm. *J. Cogn. Neurosci.* 20 (8), 1355–1370. doi:10.1162/jocn.2008.20093

Magri, G. (2014). "An Account for the Homogeneity Effect Triggered by Plural Definites and Conjunction Based on Double Strengthening," in *Pragmatics, Semantics and the Case of Scalar Implicatures*. Editor S. Pistoia Reda (London, United Kingdom: Palgrave Macmillan), 99–145.

Malamud, S. (2012). The Meaning of Plural Definites: A Decision-Theoretic Approach. *Semantics and Pragmatics* 5 (3), 1–58. doi:10.3765/sp.5.3

Marty, P., Chemla, E., and Spector, B. (2013). Interpreting Numerals and Scalar Items Under Memory Load. *Lingua* 133, 152–163. doi:10.1016/j.lingua.2013.03.006

Mayo, R., Schul, Y., and Burnstein, E. (2004). "I Am Not Guilty" vs "I Am Innocent": Successful Negation May Depend on the Schema Used for Its Encoding. *J. Exp. Soc. Psychol.* 40 (4), 433–449. doi:10.1016/j.jesp.2003.07.008

Meijer, A. M. (2016). "Referring to a Proposition 'inside' a Negated One: The Modal Particles Auch and Ook in Contextual Counterfactuals," in Proceedings of the 24th Conference of the Student Organization of Linguistics in Europe (ConSOLE). Editors K. Bellamy, E. Karvovskaya, and G. Saad (Leiden, Netherlands: University Center for Linguistics), 19–42.

Meijer, A. M., and Repp, S. (2018). "Modal Subordination of Propositional Anaphora: On the Role of Tense and the Modal Particle Ook in Contextual Counterfactuals in Dutch," in Proceedings of Sinn und Bedeutung 21:2. Editors R. Truswell, C. Cummins, C. Heycock, B. Rabern, and H. Rohde (Edinburgh: University of Edinburgh), 881–898.

Nicolae, A. C., and Sauerland, U. (2020). "Quantity Implicatures," in *The Wiley Blackwell Companion to Semantics*. Editors D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. Zimmermann (Hoboken, NJ: Wiley), 1–28.

Nieuwland, M. S., and Kuperberg, G. R. (2008). When the Truth is Not Too Hard to Handle. *Psychol. Sci.* 19 (12), 1213–1218. doi:10.1111/j.1467-9280.2008.02226.x

Nordmeyer, A. E., and Frank, M. C. (2014). The Role of Context in Young Children's Comprehension of Negation. *J. Mem. Lang.* 77, 25–39. doi:10.1016/j.jml.2014.08.002

Noveck, I. A., and Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain Lang.* 85, 203–210.

Onea, E. (2019). "Exhaustivity in It-Clefts," in *The Oxford Handbook of Experimental Semantics and Pragmatics*. Editors C. Cummins and N. Katsos (Oxford, United Kingdom: Oxford University Press), 400–417.

Orenes, I., Beltrán, D., and Santamaría, C. (2014). How Negation is Understood: Evidence From the Visual World Paradigm. *J. Mem. Lang.* 74, 36–45. doi:10.1016/j.jml.2014.04.001

Orenes, I., García-Madruga, J. A., Gómez-Veiga, I., Espino, O., and Byrne, R. M. J. (2019). The Comprehension of Counterfactual Conditionals: Evidence From Eye-Tracking in the Visual World Paradigm. *Front. Psychol.* 10, 1172. doi:10.3389/fpsyg.2019.01172

Orenes, I., Moxey, L., Scheepers, C., and Santamaría, C. (2016). Negation in Context: Evidence From the Visual World Paradigm. *Q. J. Exp. Psychol.* 69 (6), 1082–1092. doi:10.1080/17470218.2015.1063675

Papeo, L., Hochmann, J.-R., and Battelli, L. (2016). The Default Computation of Negated Meanings. *J. Cogn. Neurosci.* 28 (12), 1980–1986. doi:10.1162/jocn_a_01016

Partee, B. H. (1993). "On the 'scope of Negation' and Polarity Sensitivity," in *Functional Description of Language*. Editor E. Hajičová (Prague, Czechia: Charles University), 179–196.

Percus, O. (1997). "Prying Open the Cleft,," in Proceedings of the North East Linguistics Society 27 (NELS). Editor K. Kusumoto (Amherst, MA: GLSA), 337–351.

Potts, C., Lassiter, D., Levy, R., and Frank, M. C. (2015). Embedded Implicatures as Pragmatic Inferences under Compositional Lexical Uncertainty. *J. Semant.* 33 (4), 755–802. doi:10.1093/jos/ffv012

Rees, A., and Bott, L. (2018). The Role of Alternative Salience in the Derivation of Scalar Implicatures. *Cognition* 176, 1–14. doi:10.1016/j.cognition.2018.02.024

Repp, S. (2016). "Contrast: Dissecting an Elusive Information-Structural Notion and its Role in Grammar," in *The Oxford Handbook of Information Structure*. Editors C. Féry and S. Ishihara (Oxford, United Kingdom: Oxford University Press), 270–289.

Repp, S. (2010). Defining 'Contrast' as an Information-Structural Notion in Grammar. *Lingua* 120 (6), 1333–1345. doi:10.1016/j.lingua.2009.04.006

Repp, S., and Drenhaus, H. (2015). Intonation Influences Processing and Recall of Left-Dislocation Sentences by Indicating Topic vs. Focus Status of Dislocated Referent. *Lang. Cogn. Neurosci.* 30, 324–346. doi:10.1080/23273798.2014.923104

Roberts, C. (1996). Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics* 5 (6), 1–69. doi:10.3765/sp.5.6

Roberts, C. (2010). *Retrievability and Incomplete Descriptions*. Columbus, OH: The Ohio State University.

Rooth, M. (1992). A Theory of Focus Interpretation. *Nat. Lang. Semant.* 1 (1), 75–116. doi:10.1007/bf02342617

Rooth, M. (1985). *Association With Focus*. [dissertation]. Amherst (MA): University of Massachusetts.

Rooth, M. (1996). "Focus," in *The Handbook of Contemporary Semantic Theory*. Editor S. Lappin (London, United Kingdom: Blackwell)), 271–297.

Russell, B. (2012). Probabilistic reasoning and the computation of scalar implicatures. [dissertation]. Providence (RI): Brown University.

Ruys, E. G. (2015). On the Anaphoricity of Too. *Linguist. Inq.* 46 (2), 343–361. doi:10.1162/LING_a_00184

Santamaría, C., Espino, O., and Byrne, R. M. J. (2005). Counterfactual and Semifactual Conditionals Prime Alternative Possibilities. *J. Exp. Psychol. Learn. Mem. Cogn.* 31 (5), 1149–1154. doi:10.1037/0278-7393.31.5.1149

Santorio, P. (2018). Alternatives and Truthmakers in Conditional Semantics. *J. Philos.* 115 (10), 513–549. doi:10.5840/jphil20181151030

Sauerland, U. (2004). Scalar Implicatures in Complex Sentences. *Linguist. Philos.* 27 (3), 367–391. doi:10.1023/B:LING.0000023378.71748.db

Schulz, K. (2019). "The Similarity Approach Strikes Back: Negation in Counterfactuals," in *Proceedings of Sinn und Bedeutung 22:2*. Editors U. Sauerland, and S. Solt (Berlin, Germany: Leibniz-Centre General Linguistics), 343–360.

Schulz, K., and Van Rooij, R. (2006). Pragmatic Meaning and Non-Monotonic Reasoning: The Case of Exhaustive Interpretation. *Linguistics Philos.* 29 (2), 205–250. doi:10.1007/s10988-005-3760-4

Schwarz, F., Clifton, C., and Frazier, L. (2008). "Strengthening 'or': Effects of Focus and Downward Entailing Contexts on Scalar Implicatures," in University of Massachusetts Occasional Papers in Linguistics 33:1, UMOP 37: Semantics and Processing. Editors J. Anderssen, K. Moulton, F. Schwarz, and C. Ussery (Amherst, MA: GLSA).

Schwarzschild, R. (1993). Plurals, Presuppositions and the Sources of Distributivity. *Nat. Lang. Semant.* 2:3, 201–248. doi:10.1007/BF01256743

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27 (3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Singh, R. (2019). Context, Content, and the Occasional Costs of Implicature Computation. *Front. Psychol.* 10, 2214. doi:10.3389/fpsyg.2019.02214

Singh, R. (2008). On the Interpretation of Disjunction: Asymmetric, Incremental, and Eager for Inconsistency. *Linguist. Philos.* 31 (2), 245–260. doi:10.1007/s10988-008-9038-x

Skordos, D., and Papafragou, A. (2016). Children's Derivation of Scalar Implicatures: Alternatives and Relevance. *Cognition* 153, 6–18. doi:10.1016/j.cognition.2016.04.006

Soames, S. (1982). How Presuppositions are Inherited: A Solution to the Projection Problem. *Linguist. Inq.* 13 (3), 483–545.

Spalek, K., Gotzner, N., and Wartenburger, I. (2014). Not Only the Apples: Focus Sensitive Particles Improve Memory for Information-Structural Alternatives. *J. Mem. Lang.* 70, 68–84. doi:10.1016/j.jml.2013.09.001

Spychalska, M., Kontinen, J., and Werning, M. (2016). Investigating Scalar Implicatures in a Truth-Value Judgement Task: Evidence From Event-Related Brain Potentials. *Lang. Cogn. Neurosci.* 31 (6), 817–840. doi:10.1080/23273798.2016.1161806

Stalnaker, R. (1968). "A Theory of Conditionals," in *Studies in Logical Theory: Essays*. Editor J. Cornman (Oxford: Blackwell), 98–112.

Tessler, M. H., and Goodman, N. D. (2019). The Language of Generalization. *Psychol. Rev.* 126 (3), 395–436. doi:10.1037/rev0000142

Thompson, V. A., and Byrne, R. M. J. (2002). Reasoning Counterfactually: Making Inferences About Things that Didn't Happen. *J. Exp. Psychol. Learn. Mem. Cogn.* 28 (6), 1154–1170. doi:10.1037/0278-7393.28.6.1154

Tian, Y., and Breheny, R. (2016). "Dynamic Pragmatic View of Negation Processing," in *Negation and Polarity: Experimental Perspectives*. Editors P. Larrivée and C. Lee (Cham, Switzerland: Springer), 21–43.

Tian, Y., Breheny, R., and Ferguson, H. J. (2010). Why we Simulate Negated Information: A Dynamic Pragmatic Account. *Q. J. Exp. Psychol.* 63 (12), 2305–2312. doi:10.1080/17470218.2010.525712

Tian, Y., and Breheny, R. (2019). "Negation," in *The Oxford Handbook of Experimental Semantics and Pragmatics*. Editors C. Cummins and N. Katsos (Oxford, United Kingdom: Oxford University Press), 194–207.

Tian, Y., Ferguson, H., and Breheny, R. (2016). Processing Negation Without Context - Why and When we Represent the Positive Argument. *Lang. Cogn. Neurosci.* 31 (5), 683–698. doi:10.1080/23273798.2016.1140214

Tieu, L., Križ, M., and Chemla, E. (2019). Children's Acquisition of Homogeneity in Plural Definite Descriptions. *Front. Psychol.* 10, 2329. doi:10.3389/fpsyg.2019.02329

Tjuka, A., Nguyen, H. T. T., and Spalek, K. (2020). Foxes, Deer, and Hedgehogs: The Recall of Focus Alternatives in Vietnamese. *Lab. Phonol.* 11 (1), 16. doi:10.5334/labphon.253

Tomioka, S. (2021). Scalar Implicature, Hurford's Constraint, Contrastiveness and How They All Come Together. *Front. Commun.* 5, 133. doi:10.3389/fcomm.2020.461553

Tomlinson, J. M., Bailey, T. M., and Bott, L. (2013). Possibly All of That and Then Some: Scalar Implicatures are Understood in Two Steps. *J. Mem. Lang.* 69 (1), 18–35. doi:10.1016/j.jml.2013.02.003

Tonhauser, J., Beaver, D., Roberts, C., and Simons, M. (2013). Toward a Taxonomy of Projective Content. *Language* 89 (1), 66–109. doi:10.1353/lan.2013.0001

Trinh, T. (2019). Exhaustification and Contextual Restriction. *Front. Commun.* 4, 47. doi:10.3389/fcomm.2019.00047

Trinh, T., and Haida, A. (2015). Constraining the Derivation of Alternatives. *Nat. Lang. Semant.* 23 (4), 249–270. doi:10.1007/s11050-015-9115-y

Van Rooij, R., and Schulz, K. (2004). Exhaustive Interpretation of Complex Sentences. *J. Log. Lang. Inf.* 13 (4), 491–519. doi:10.1007/s10849-004-2118-6

Van Rooij, R. (2004). Utility, Informativity and Protocols. *J. Phil. Log.* 33 (4), 389–419. doi:10.1023/B:LOGI.0000036830.62877.ee

van Tiel, B., and Schaeken, W. (2017). Processing Conversational Implicatures: Alternatives and Counterfactual Reasoning. *Cogn. Sci.* 41, 1119–1154. doi:10.1111/cogs.12362

Velleman, D., Beaver, D., Destruel, E., Bumford, D., Onea, E., and Coppock, L. (2012). "It-Clefts are IT (Inquiry Terminating) Constructions,," in Proceedings of Semantics and Linguistic Theory. Editor A. Chereches (Chicago, IL: Linguistic Society of America), 441–460.

von Stutterheim, C. (1997). *Einige Prinzipien des Textaufbaus*. Tübingen, Germany: Niemeyer.

Wagner, M. (2012). Contrastive Topics Decomposed. *Semant. Pragmat.* 5 (8), 1–54. doi:10.3765/sp.5.8

Wason, P. C. (1972). In Real Life Negatives are False. *Log. Anal.* 15 (57/58), 17–38.

Watson, J. M. (1979). Referential description by children in negative form. *British J. Psychol.* 70, 199–204.

Wason, P. C., and Jones, S. (1963). Negatives: Denotation and Connotation. *Br. J. Psychol.* 54 (4), 299–307. doi:10.1111/j.2044-8295.1963.tb00885.x

Wason, P. C. (1961). Response to Affirmative and Negative Binary Statements. *Br. J. Psychol.* 52 (2), 133–142. doi:10.1111/j.2044-8295.1961.tb00775.x

Wason, P. C. (1965). The Contexts of Plausible Denial. *J. Verb. Learning Verb. Behav.* 4 (1), 7–11. doi:10.1016/S0022-5371(65)80060-3

Watson, D. G., Tanenhaus, M. K., and Gunlogson, C. A. (2008). Interpreting Pitch Accents in Online Comprehension: H* vs. L+H*. *Cogn. Sci.* 32 (7), 1232–1244. doi:10.1080/03640210802138755

Weber, A., Braun, B., and Crocker, M. W. (2006). Finding Referents in Time: Eye-Tracking Evidence for the Role of Contrastive Accents. *Lang. Speech* 49 (3), 367–392. doi:10.1177/00238309060490030301

Willer, M. (2018). Simplifying With Free Choice. *Topoi* 37 (3), 379–392. doi:10.1007/s11245-016-9437-5

Wiswede, D., Koranyi, N., Müller, F., Langner, O., and Rothermund, K. (2013). Validating the Truth of Propositions: Behavioral and ERP Indicators of Truth Evaluation Processes. *Soc. Cogn. Affect. Neurosci.* 8, 647–653. doi:10.1093/scan/nss042

Yan, M., and Calhoun, S. (2019). Priming Effects of Focus in Mandarin Chinese. *Front. Psychol.* 10, 1985. doi:10.3389/fpsyg.2019.01985

Zimmermann, M. (2008). "Contrastive focus and emphasis," in *Acta Linguistica Hungarica is a Quarterly Hungarian Peer-reviewed Academic Journal in the Field of Linguistics*. Editor F. Kiefer (Akadémiai Kiadó), 347–360.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 April 2019
doi: 10.3389/fpsyg.2019.00877

Check for updates

Asymmetries Between Direct and Indirect Scalar Implicatures in Second Language Acquisition

Shuo Feng and Jacee Cho*

Department of English, University of Wisconsin-Madison, Madison, WI, United States

A direct scalar implicature (DSI) arises when a sentence with a weaker term like *sometimes* implies the negation of the stronger alternative *always* (e.g., John sometimes (∼ not always) drinks coffee). A reverse implicature, often referred to as indirect scalar implicature (ISI), arises when the stronger term is under negation and implicates the weaker alternative (e.g., John doesn't always (∼ sometimes) drink coffee). Recent research suggests that English-speaking adults and children behave differently in interpreting these two types of SI (Cremers and Chemla, 2014; Bill et al., 2016). However, little attention has been paid to how these two types of SI are processed in a non-native, or second language (L2). By using a covered box paradigm, this study examines how these two types of SI are computed and suspended in a second language by measuring the visible vs. covered picture selection percentage as well as response times (RTs) taken for the selection. Data collected from 26 native speakers of English to 24 L1-Chinese L2-English learners showed that unlike native speakers, L2 speakers showed asymmetries in their generation and suspension of DSI and ISI. That is, L2 speakers computed DSI more often than ISI, but they suspended ISI more frequently than DSI. Furthermore, our RT data suggested that L2 speakers suspended ISI not only more frequently but also significantly faster than DSI. Regarding the asymmetrical behavior among L2 speakers, we consider the number of alternative meanings involved in DSI vs. ISI suspension and different routes to the suspension of SI.

Keywords: direct and indirect scalar implicatures, alternatives, SI suspension, second language acquisition, covered-box paradigm

OPEN ACCESS

Edited by:
Katharina Spalek, Humboldt University of Berlin, Germany

Reviewed by:
Chris Cummins, The University of Edinburgh, United Kingdom
David Miller, University of Reading, United Kingdom

***Correspondence:**
Shuo Feng
sfeng27@wisc.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *29 November 2018*
Accepted: *03 April 2019*
Published: *24 April 2019*

Citation:
Feng S and Cho J (2019) Asymmetries Between Direct and Indirect Scalar Implicatures in Second Language Acquisition. Front. Psychol. 10:877. doi: 10.3389/fpsyg.2019.00877

INTRODUCTION

Many linguistic forms are interpreted semantically and pragmatically, which generate more than one meaning from the same form. This forces the hearer to consider all the alternative meanings and choose the meaning that is most appropriate in a given context. Alternative meanings are argued to be accessed and computed separately from the semantic meaning (Rooth, 1985, 1992, 2016). For example, the utterance in (1a) has the semantics of (1b) but implicates the proposition in (1c). Similarly, (2a) can be interpreted semantically as in (2b) and also pragmatically as in (2c).

(The symbol "∼" in this paper is used to indicate implied meaning).

(1) a. Bob sometimes went to school (DSI).
b. Bob went to school at least once and possibly all the time (always).
c. ∼ Bob didn't always go to school.
(2) a. Bob did not always go to school (ISI).
b. Bob failed to go to school at least once and possibly never went to school.
c. ∼ Bob sometimes went to school.

The linguistic phenomenon that involves a set of alternatives in terms of informational strength (e.g., < *never, rarely, sometimes, often, always* >, < *some, most, all* >) is called scalar implicature (SI). Generating an implicature from a weaker term like *sometimes* by negating the stronger alternative *always*, as in (1a) and (1c), is often referred to as *direct scalar implicature* (DSI). An implicature derived from the stronger term under negation by considering the weaker alternative like (2a) and (2c) is called *indirect scalar implicature* (ISI).

An account for *why* and *how* we make inferences like (1c) and (2c) beyond what was said in (1a) and (2a) comes from the philosopher Grice's (1975) theory of inferential communication. According to the theory, we conduct our communication based on rational expectations and principles to meet the goals of communication. He called these principles and expectations 'maxims'. One of the maxims, the *Quantity Maxim,* states that interlocutors are cooperative by making their contribution as informative as is required but no more informative than is required. On this account, saying *Bob sometimes went to school* while he always went to school is true but underinformative, thus violating the Quantity Maxim. This prompts the hearer to make the inference that the stronger term *always* does not hold since the speaker would have said *Bob always went to school* following the Quantity Maxim.

Drawing on Grice's theory of inferential communication, Levinson (2000) proposes a Default Inference account of scalar implicatures to explain how scalar inference arises in real-time communications. According to Levinson (2000), scalar implicatures are default inferences that are generated automatically and are canceled only when the context calls for it. Scale terms such as *sometimes* are stored in our memory in association with alternative terms like *always*, *often*, and *rarely* due to habitual generation of the implicatures for *sometimes* (i.e., 'not always') in everyday communications (Gazdar, 1979; Levinson, 1983, 2000). Since scalar implicatures are made by default, they require little cognitive efforts from a processing point of view. Some recent psycholinguistic studies on adult native speakers provided evidence for the Default account (Grodner et al., 2010; Lewis and Phillips, 2011).

Arguing against the default view is a context-driven view such as the Relevance Theory supported by Sperber and Wilson (1986/1995) and Carston (2004). Within this approach, utterances are enriched with inferences only if they are relevant to reach the speaker's intended meaning in a given context. From the point of view of the Relevance Theory, the implicated meaning of *sometimes* (∼ *not always*) in (1c) or the implied meaning of *not always* (∼ *sometimes*) in (2c) are not derived automatically by default, but rather are generated effortfully by canceling the initial literal meaning. In short, the context-driven approach argues that mental effort is required to derive contextual effects to generate scalar implicatures. As a matter of fact, a growing number of recent psycholinguistic studies on native speakers indicate that scalar implicature involves an extra cognitive process evidenced by slower response times in sentence judgment tasks (Bott and Noveck, 2004), longer reading times in self-paced reading tasks (Breheny et al., 2006; Bergen and Grodner, 2012), and delayed eye fixations in a visual world eye-tracking task (Huang and Snedeker, 2009, 2011). For example, Bott and Noveck (2004) examined the generation of SI in adult native speakers of French by measuring response times (RTs) in a sentence-verification task containing underinformative (i.e., pragmatically infelicitous) sentences like (3a). Such underinformative sentences are false with a scalar inference (*some but not all* in (3b)) and true without the inference (*some and possibly all* as in (3c)). Therefore, if participants compute SI (*some but not all*), they would answer 'False' to the statement in (3a) because all elephants are mammals. If participants answer 'True', it means that participants suspend SI inference and interpret *some* as *some and possibly all* as in (3c).

(3) a. Some elephants are mammals.
b. ∼ Not all elephants are mammals.
c. Possibly all elephants are mammals.

Additionally, to investigate the speed of responses, participants in the experiment[1] were asked to judge such a sentence under two different instructions. Under the 'Logical' condition, participants were instructed to interpret *some* as *some and possibly all* whereas under the 'Pragmatic' condition, participants were instructed to interpret *some* as *some but not all.*

The results supported the Relevance Theory account. That is, when participants were asked to judge pragmatically, they spent more time in evaluating the underinformative sentences than when they were under the Logical condition. It further indicated that maintaining the SI inference was not effortless in processing and SI computation required extra cognitive effort, as evidenced in longer RTs. This finding was also confirmed by subsequent studies using various methodologies (Degen and Tanenhaus, 2011; Bott et al., 2012).

By employing event-related potentials (ERP) techniques, a large number of studies have investigated the integration of semantic interpretation and pragmatic inference of sentence processing. Noveck and Posada (2003) suggested a smaller N400 effect in underinformative sentences than both semantically and pragmatically acceptable sentences. However, Nieuwland et al. (2010, Experiment 1) reported a similar pattern of N400 in reading underinformative sentences only among participants with low pragmatic ability. By using a picture-sentence verification methodology, Politzer-Ahles et al. (2013) tested Mandarin Chinese speakers' interpretation of the Chinese scalar item *you de* 'some of' in underinformative sentences.

[1] Bott and Noveck (2004) conducted four experiments. In this paper, we limit our discussion to the first experiment.

The ERP results showed a sustained negativity effect when the pragmatic interpretation of scalar items was not consistent with the context, indicating that suspending pragmatic meaning and activating semantic meaning required extra cognitive effort. More importantly, the authors found a qualitatively different ERP pattern of Chinese scalar items in semantically infelicitous sentences compared to pragmatically infelicitous sentences. It indicates that the reanalysis process of canceling or suspending the pragmatic interpretation is distinctively different from the process of accessing the semantic meaning.

It has been suggested that canceling SI may require additional cognitive efforts (Bill et al., 2015). Being an inference, not linguistically encoded meaning, scalar implicatures can be explicitly canceled without logical contraction. For example, in (4), the inference *not always* of the DSI item *sometimes* is explicitly canceled in Speaker B's utterance. Similarly, the inference *sometimes* of the ISI item *not always* in (5) is obviously absent in Speaker B's utterance.

(4) A: Bob was very sick last week. But he sometimes went to school last week.
B: Yes, in fact, he always went to school last week.
(5) A: Bob was very sick last week. So, he didn't always go to school last week.
B: Yes, in fact, he never went to school last week.

There are two routes to the no-inference interpretation. The first route is the following. Under the assumption that a literal meaning without SI is default as proposed by the Relevance Theory, a no-inference reading can be done simply by not generating SI. This way of computing a no-inference interpretation is argued to be cognitively less demanding than generating SI since no-inference is the default interpretation. This is why young children, unlike adults, often prefer literal, no-inference interpretations for scalar items (Smith, 1980; Chierchia et al., 2001; Noveck, 2001; Papafragou and Musolino, 2003). The second route to the no-inference interpretation is to cancel SI after it has been generated first. Whether one's no-inference interpretation is computed through the first route (i.e., not generating SI at all) or through the second route (i.e., canceling SI) can be teased apart via measuring and comparing response times. We will return to this issue in the methodology section. In this paper, the term *SI suspension* is used generally to refer to the no-inference reading achieved either by not generating SI (in young children's case) or by canceling SI via re-calculation.

Traditionally, DSIs and ISIs are considered to be the same type of inference; thus, it was assumed that they are involved in the same mechanisms and similar processing efforts. However, recent studies have shown that adults and children behave differently between DSIs and ISIs. One proposal made by Spector (2007) and Chierchia et al. (2012) is that ISIs are obligatory implicatures while DSIs are non-obligatory. According to this proposal, generating DSIs should require more efforts than generating ISIs, but suspending obligatory ISIs should be harder than suspending non-obligatory DSIs. That is, since ISIs are obligatory, interpretations with ISIs (not always ∼ 'sometimes') should be easier to process than interpretations without ISIs (not always ∼ never). These two approaches make different predictions about how DSIs and ISIs are generated and processed.[2]

To test whether DSI and ISI are the same kind of inference, Cremers and Chemla (2014) examined the generation of ISI and compared with DSI using a sentence verification task. In their second experiment, participants were asked to judge whether sentences with ISI inference like (6) are true or false against a cover story.

(6) Not all of the [land animals] were fortified.

All sentences were expected to be true under the logical reading (*not all and none*) by suspending the inference but false under the pragmatic reading (*not all but some*). In addition, participants also received explicit instruction on how to interpret these sentences. Half of the participants were assigned to the No-SI group (equivalent to the Logical condition in Bott and Noveck, 2004) and the other half belonged to the SI group (equivalent to the Pragmatic condition). The findings suggested that ISI computation was cognitively more demanding and further indicated a general uniformity for the mechanism that gives rise to both DSI and ISI: scalar implicatures are associated with a delay regardless of the type of SI.

While DSI and ISI seem to be generated in a similar way, their suspension appears to be done through different mechanisms or require varying degree of cognitive efforts as shown in Bill et al. (2016). Instead of using the truth-value judgment paradigm like a sentence verification task, Bill et al. (2016) employed a covered box method developed by Huang et al. (2013). The covered box paradigm differs from the truth-value judgment methodology in that it explicitly offers the non-dominant no-inference interpretation, which encourages participants to consider both inference and no-inference interpretations for the test sentence. That is, while the truth value judgment paradigm is good for examining inference computation, the covered box method is well suited to an investigation of inference suspension.

Using the covered box method, Bill et al. (2016) examined and compared three types of inference: presupposition, DSI and ISI. However, we limit our attention here to Bill et al.'s comparison of DSI and ISI since discussion of presuppositions falls outside the scope of our paper. In Bill et al. (2016), participants were given a test sentence with a visible picture and a black covered box. They were asked to choose the visible picture if it matches the test sentence and choose the covered box if the visible picture does not match the test sentence. Example trials of DSI and ISI conditions are provided in **Figures 1**, **2**, respectively (**Figures 1**, **2** are adapted from Bill et al., 2016).

[2]There are other proposals regarding the nature of DSI, ISI and Presupposition. For example, Chemla (2009) and Romoli (2012) argue that ISIs are presuppositions. The ISI as Presupposition approach states that the stronger scalar terms presuppose their weakest competitor. The statement "John always drinks coffee" presupposes that "John sometimes drinks coffee." Since presuppositions project under negation, "John doesn't always drink coffee" should also presuppose "John sometimes drinks coffee." In this approach, "sometimes" is not an implicature but a presupposition of "not always". Since our study deals only with DSIs and ISIs, we will not consider the ISI as presupposition approach in discussing our data.

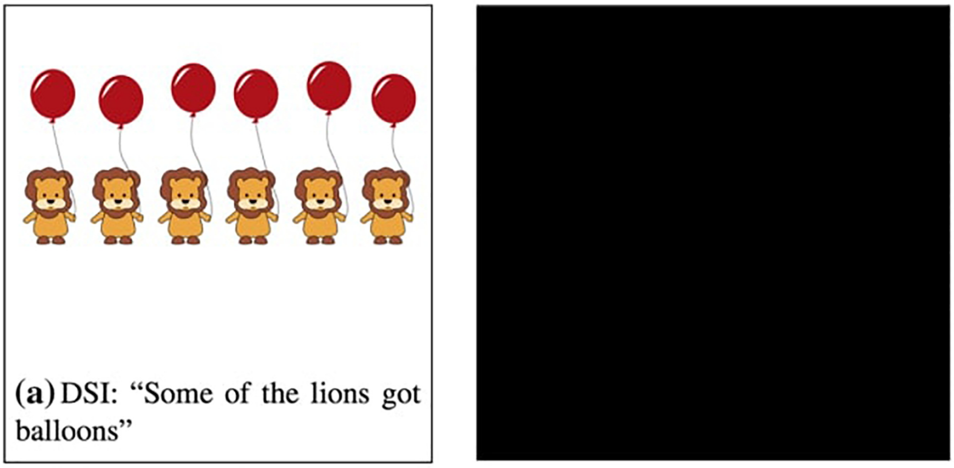


FIGURE 1 | Example trial of the DSI condition [adapted from Bill et al. (2016)].

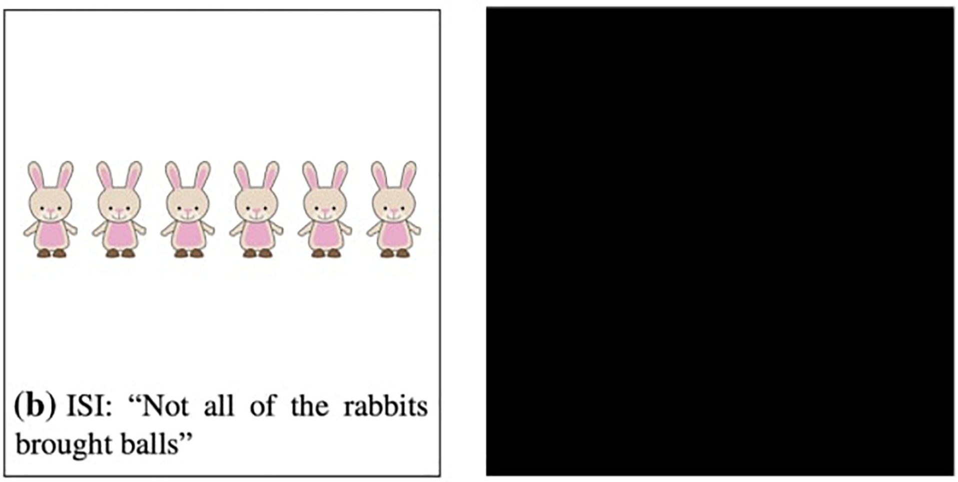


FIGURE 2 | Example trial of the ISI condition [adapted from Bill et al. (2016)].

The selection of a covered box in each condition indicates the generation of SI whereas the selection of a visible picture suggests the suspension of SI. For instance, in **Figure 1**, the visible picture depicts a no-inference reading of *some lions*, i.e., *some and possibly all lions*, thus selecting the visible picture indicates suspension of DSI. If participants compute DSI (*some but not all*), they would reject the visible no-inference reading and select the covered box. In **Figure 2**, the visible picture shows a no-inference reading of *not all*, i.e., *none of the rabbits*. Selecting the visible pictures indicates ISI suspension and choosing the covered-box suggests ISI computation.

There were three groups of English-speaking participants: adults, 4—5 year olds, and 7 year olds. Results showed that adults generated DSI significantly more often than ISI whereas 4-5 year olds and 7 year olds computed ISI significantly more frequently than DSI. Adults were more likely to suspend the inference in ISI than in DSI (a low percentage of selecting covered-box in ISI vs. a high percentage in DSI), while children were more likely to suspend the inference in DSI than in ISI (the opposite percentage pattern to adults).

In sum, there is a general uniformity of processing behavior between DSI and ISI computation such that DSI and ISI are computed at similar rates. However, there are asymmetries between DSI and ISI suspension. English-speaking adults are more likely to suspend ISI than DSI whereas children are more likely to suspend DSI than ISI.

Understanding how DSIs and ISIs are computed and suspended is important not only in linguistic and psycholinguistic theory but also in L2 acquisition theory. Previous research into SI in L2 acquisition has shown that SI computation is not a problem for L2 speakers. In fact, L2 speakers tend to generate SIs more than native speakers do (Lieberman, 2009; Slabakova, 2010; Miller et al., 2016; Snape and Hosoi, 2018). Slabakova (2010) hypothesizes that L2 speakers compute SI more than native speakers because SI cancelation may present challenges to L2 speakers. This issue, however, has not been tested empirically. The present study aims to test whether differences between native speakers and L2 speakers lie in SI suspension rather than SI computation using the covered box paradigm (the logic of this method will be discussed in the next section). Moreover, while there is an increasing number of L2 studies on DSI, little research has been done on ISI in L2 acquisition. To fill this gap, this study examines and compares computation and suspension of DSI vs. ISI by focusing on scalar items like < *sometimes, always* > . Thus, findings of this study would advance our understanding of how alternative meanings are considered in the generation or suspension of SI in an L2.

SCALAR IMPLICATURES IN ADULT L2 SPEAKERS

The experimental work on the inference computation in adult L2 learners is rather limited. The first study is Slabakova's (2010) study on how L1-Korean L2-English learners process scalar expressions, such as quantifiers *some* and *all* in their L1 Korean vs. L2 English. The critical experimental item on *some* is (7), which is logically true but pragmatically infelicitous. If participants reject such sentences, it provides clear evidence that participants are able to derive SI and compute the pragmatic reading of *some* as *some but not all*. Acceptance of these sentences indicates that participants suspend SI and generate the logical meaning of *some* as *some and possibly all.*

(7) Some elephants have trunks.
(Slabakova, 2010, p. 2452)

The results showed that Korean learners of English successfully acquired scalar implicatures in their L2. However, differences in response patterns still existed between native speakers and learners. That is, L1-Korean learners of L2-English were more likely than monolingual English or Korean speakers to reject pragmatically infelicitous sentences like (7). One possible explanation proposed by Slabakova (2010) is that it is easier to conjure up situations to make underinformative sentences plausible. For example, if one can think of a situation where some elephants' trunks got cut due to accidents, the sentence in (7) is felicitous. Another possibility is differential ability to SI suspension. That is, if one cancels the [not all] implicature, the statement in (7) should be interpreted as 'At least one and possibly all elephants have trunks', which is true. Since SI suspension arguably requires more cognitive efforts, it might be more difficult to do in an L2 under the assumption that less cognitive resources are available for L2 processing than L1 processing (Green, 1986, 1998)[3].

A similar study was carried out on L1-English L2-Spanish learners' interpretation of Spanish quantifiers (Miller et al., 2016) and potential L1 influence in this domain. Unlike Korean that has only one lexical item roughly equal to the English scalar term *some*, Spanish has two: *algunos* and *unos.* While both words have the pragmatic interpretation *some but not all*, only *unos* has the additional logical interpretation *some and possibly all.* With an inherent partitive feature, *algunos* cannot be inferred logically. Imagine a situation that someone has four dogs. When a postman arrives, three out of four dogs barked at the postman in front of the door. In this situation, using either *algunos* or *unos* to mean *some but not all* is felicitous in *Some dogs barked at the postman.* If all the four dogs barked at the postman, the logical interpretation is desired. Thus, it is only felicitous to use *unos,* as in (8a), but infelicitous to use *algunos,* as in (8b).

(8) Context – All four dogs bark at the postman.
a. *Unos perros ladraron al cartero.*
"Some dogs barked at the postman."
b. **Algunos perros ladraron al cartero.*
"Some dogs barked at the postman."
(Miller et al., 2016, p. 131)

The fact that Spanish and English do not have a one-to-one mapping on *some* may present further challenges to L2 learners of Spanish. Miller et al. (2016) tested L1-English L2-Spanish learners' acquisition of the two Spanish scalar terms *algunos* and *unos* through a truth-value video acceptability judgment task. They discovered that English learners were able to obtain a native-like judgment on the two Spanish scalar terms irrespective of the fact that English has a different scalar implicature system. Specifically, not replying on a 1:1 mapping between English and Spanish scalar terms, English learners were less likely to accept *algunos* in non-partitive contexts but were equally likely to accept *unos* despite partitive or non-partitive contexts.

Similar findings were obtained in Snape and Hosoi's (2018) study on L1-Japanese speakers' interpretation of *some* in L2 English. The Japanese quantifier *ikutsuka* translates into *some* in English, as in (9).

(9)	Akai	maru no	naka	ni	banana ga
	red	circle-POSS	inside	of	banana-NOM
	ikutsuka	arimasu	ka		
	some	to be	Q		

'Are some bananas in the red circle?'

However, unlike English *some* (or Spanish *algunos*), Japanese *ikutsuka* does not have a partitive meaning (*not all*), that is, it does not implicate the *some but not all* meaning. Using a picture-based acceptability judgment task, Snape and Hosoi (2018) examined whether intermediate-level L1-Japanese L2-English learners overaccept pragmatically infelicitous sentences due to L1 transfer and whether such L1 influence would disappear as the proficiency level increases. Conforming to previous studies, Snape and Hosoi (2018) found that L1-Japanese speakers had no

[3]The assumption that L2 grammars have limited cognitive resources is based on generally accepted conclusions from psycholinguistic research into bilingualism that both languages are active at all times in the mind of bilinguals (Marian and Spivey, 2003; Dijkstra, 2005; Kroll et al., 2006; Kroll et al., 2015) and having two languages in one mind is cognitively costly because suppressing one language during performance of the other requires cognitive effort (see Green, 1986, 1998 for discussion on inhibitory control/linguistic inhibition in bilingual language performance).

difficulty in deriving scalar implicatures despite the mismatches between L2-English *some* and L1-Japanese *ikutsuka* 'some'. Moreover, there was no proficiency effect.

Lin (2016), employing a series of real-time psycholinguistic experiments on Chinese learners' acquisition of *some,* contributed to knowledge of L2 speakers' processing mechanism of scalar implicatures. The first experiment used a Truth Value Judgment task. After reading a context sentence "John has many dictionaries. Some of the dictionaries are used", participants were asked to judge whether the following target sentences were true or false: "Some and possibly all of the dictionaries are used" or "Some but not all of the dictionaries are used." Results of the first experiment showed that it was faster for Chinese speakers to compute the pragmatic interpretation of *some* as *some but not all* and it took them more time on rejecting this interpretation. When suspending SI and generating the logical reading (*some and possibly all*), Chinese participants spent almost twice as much time as they did in responding to the pragmatic interpretation. Additionally, they were more likely to reject the logical interpretation. The findings were in line with previous experimental results that adults favor the pragmatic interpretation where the SI inference is present.

The second experiment in Lin (2016) was motivated by the fact that when participants were given unlimited time to respond, they were able to come up with an alternative plausible situation that would fit the sentence at hand. In order to prevent additional brainstorming, in the second experiment, participants were required to respond within a certain amount of time. What is interesting about the finding was when Chinese speakers were pressed for time, the rejection rate of the logical interpretation *some and possibly all* was noticeably increased. In other words, Chinese participants were more likely to reject the suspension of SI when they were under the time pressure. This revealed that suspending scalar items (the logical interpretation) required more cognitive capacity and when L2 speakers' processing capacity was artificially constrained (e.g., when they were pressed for time), they preferred the cognitively less demanding reading (the pragmatic reading) by computing SI.

In brief, L2 research on SI has shown that generating SI inference is not difficult for L2 speakers and suggests that suspending SI inference may be challenging to L2 speakers. However, the methodology used in previous L2 studies could not tease apart whether differences between L1 and L2 speakers in their rate of SI interpretation is due to difficulties associated with SI suspension in an L2. Our study aims to examine this issue through an investigation of L2 learners' time course of generating and suspending DSIs and ISIs by employing the covered box paradigm. In this study, we focus on only one type of scalar expressions, namely frequency adverbs like < *never, sometimes, always* >.

THE PRESENT STUDY

Research Questions

In light of prior research on DSI and ISI, the present study addresses the following research questions:

RQ1: Do native and L2 speakers differ in *generating* DSI or ISI?
RQ2: Do native and L2 speakers differ in *suspending* DSI or ISI?

Methodology

The method used in this experiment was the *covered box paradigm* (Huang et al., 2013), as discussed in Section 2. This paradigm has been successfully applied to explore implicatures (Huang et al., 2013) and presuppositions (Schwarz, 2014; Zehr et al., 2016; Romoli and Schwarz, 2015), especially regarding suspension of an inference. Compared to a traditional picture-selecting task, the difference with the *covered box paradigm* is that it includes a covered box (see the invisible or the hidden picture on the right in **Figure 3**). Participants were told that there is one picture hidden under the black box. In the current experiment, the instruction on the *covered box paradigm* was if the visible picture matches the stimuli, participants should choose the visible picture. If the visible picture does not match the stimuli, the match must be under the black box and participants should choose the covered box. The advantage of using a covered box is that it is "...useful for testing for the availability of non-dominant interpretations..." (Romoli and Schwarz, 2015, p. 225). The non-dominant interpretation, or the suspension of an inference, is the No-inference visible meaning where the SI inference is absent in the current study. By employing

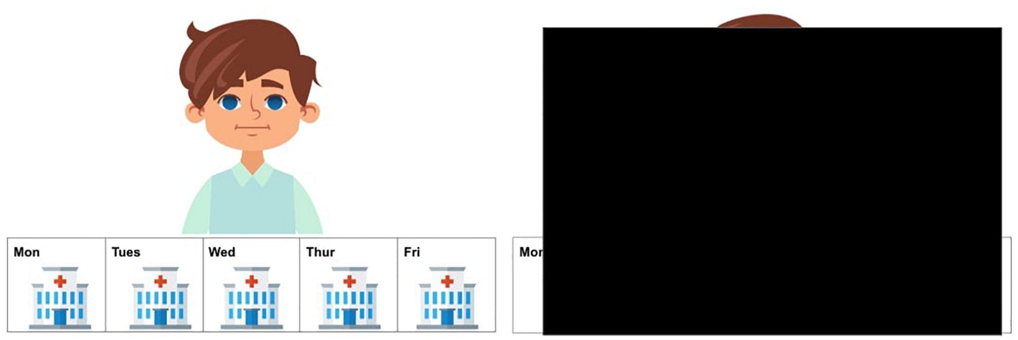


FIGURE 3 | A test trial of the stimulus *Thomas sometimes went to the hospital last week*.

the covered-box paradigm, the SI suspension reading can be displayed explicitly through a visible picture and participants are forced to consider whether the shown picture corresponds to the stimulus. A rejection of the No-inference visible picture (instead choosing the covered box) clearly indicates that the SI suspension or no-inference interpretation is not available to the participants. The same rationale also applies to the dominant interpretation (the Inference visible meaning in the present study). The visible picture in **Figure 3** displays a suspension, or No-inference interpretation which is not compatible with an Inference reading that the implicature is present, *Thomas didn't always go to the hospital.*

Test Design

In this experiment, two factors were manipulated in a 2x2 design: SI type and Visible picture. The SI type factor has two levels which are the two kinds of SI we discussed, DSI and ISI. The Visible picture factor has two levels, depending on whether the visible picture shows the SI inference (Inference) or does not display the inference (No-inference). These two factors were crossed to create four conditions: (i) DSI with a visible picture depicting the inference in (10b), (ii) DSI with a visible picture depicting a no-inference reading, like (10c), (iii) ISI with a visible picture depicting the inference in (11b), and (iv) ISI with a visible picture depicting a no-inference reading, as in (11c).

(10) a. DSI: Thomas sometimes went to school last week.
b. Inference: ∼ Thomas didn't always go to school last week.
c. No-inference: Thomas always went to school last week.

(11) a. ISI: Thomas didn't always go to school last week.
b. Inference: ∼ Thomas sometimes went to school last week.
c. No-inference: Thomas never went to school last week.

To convert (10b-c) and (11b-c) into visual stimuli to fit the *covered box paradigm,* the 5-day calendar-strip design was adapted which has been commonly used commonly to investigate the availability of presupposition interpretations was adapted for our study (Schwarz, 2014; Bill et al., 2015; Romoli and Schwarz, 2015; Bacovcin et al., 2016). In this experiment, the calendar-strip contains icons of various activities and locations from Monday to Friday[4]. A continuous appearance of an activity or a location means that this action has been repeated everyday whereas a mixture of activities or locations indicates that the first action has been stopped at some point and a new action has started[5]. **Table 1** displays four sample visible pictures for the four target conditions.[6] The two Inference pictures (12–13) were consistent with a SI interpretation, as in (10b) and (11b). The two No-inference pictures (14-15) illustrated (10c) and (11c) where the icon of hospital in (14) and circus in (15) was shown from Monday to Friday, blocking the SI interpretation. Half of the visible pictures of DSI and ISI were in the Inference condition and were predicted to be selected by both native and L2 speakers, given the preference of the inference or pragmatic interpretation of scalar items in the literature. The other half of the visible pictures were in the No-inference condition and, based on suspension or computation of SI, different response behavior was predicted. Selecting the No-inference visible picture indicates suspension of the SI inference whereas rejecting the No-inference visible picture (instead selecting the covered box) suggests the computation of SI.

In addition to target conditions, we also included controls and fillers, using the same *covered box* method. Half of the visible pictures of controls and fillers matched the stimuli and the other half did not, calling for the selection of the covered box. Controls were used to check if participants understood the task correctly and the sentence stimuli were simple negated and affirmative sentences. For instance, in **Table 2**, the visible picture of *Louis went to the train station on Wednesday and Friday* had a train station icon on Wednesday and Friday and thus triggered the visible picture selection. The visible picture for *Edward didn't go to the movies on Thursday and Friday* had a movie icon on Thursday and Friday and participants were expected to choose the covered box.

Two types of fillers were included in this experiment. The first type was created using a presupposition trigger *stop* in both affirmative and negated sentences, e.g., *Thomas stopped going to the hospital on Wednesday* and *Bob didn't stop going to school on Wednesday.* The second type of fillers had *again*, such as *Phoebe went to the gym again on Wednesday during the week.*

Procedure and Participants

Twenty-six native English speakers and twenty-four L1-Chinese L2-English learners participated in this study and they were students at a Midwest University in the United States. After signing consent forms[7], all participants finished three tasks: a brief background questionnaire, a proficiency test and a covered-box task. The background questionnaire collected participants' information about gender, age and years of studying English. The proficiency test was based on the Common European Framework of Reference for Languages (CEFR) containing 40 items with a maximum score of 40. The summary of participants' information is shown in **Table 3**.

All participants completed the covered-box task on a computer where the program E-prime was used to display stimuli and collect data. The choice of pictures was achieved by clicking

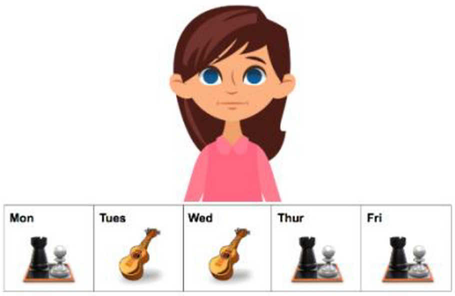
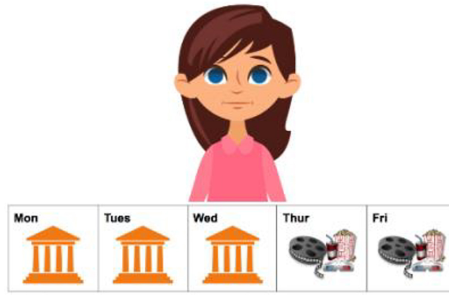
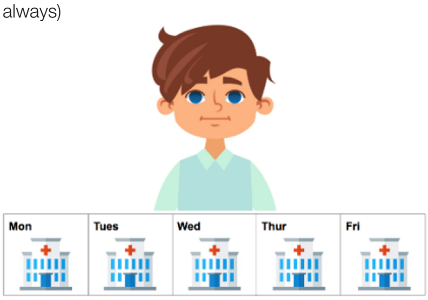
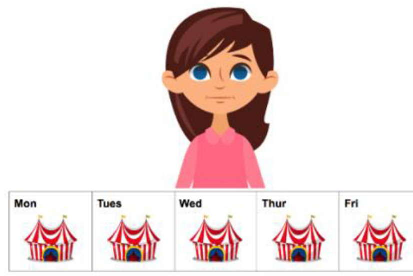
[4]A reviewer commented that the calendar strip does not include Saturday and Sunday which could leave room for participants to wonder whether the character might do something over the weekend. We provided our participants with the instruction that the calendar shows the character's activities last week from Monday to Friday. Sat and Sun were not in the scope of consideration. However, we acknowledge that the 5-day calendar might have triggered some participants to have a partitive meaning.

[5]Participants were told that the icon for one day represented that the character went to the place only or did that one activity only. For instance, if a hospital icon appears on Monday, it means that the character only went to the hospital on Monday, nowhere else.

[6]As pointed out by a reviewer, activities described in test items vary considerably from going to school to playing guitar. While 'always going to school' entails 'going to school every day', 'always playing guitar' may entail 'playing guitar every day and all day long', that is, playing guitar every day for 5 minutes would not be described as 'always playing guitar'. We acknowledge that this methodological issue could possibly influence participants' interpretation.

[7]Participants were all above 18 years old and gave written informed consent.

TABLE 1 | Four test conditions in a 2x2 factorial design: SI type (DSI vs. ISI) and Visible picture (Inference vs. No-inference).

SI type / Visible picture	DSI: *sometimes*	ISI: *not always*
Inference	(12) Daisy sometimes played guitar last week. (~ not always) 	(13) Raquel didn't always go to the movies last week. (~ sometimes) 
No- Inference	(14) Bobby sometimes went to the hospital last week. (~ always) 	(15) Lily didn't always go to the beach last week. (~ not even once) 

For ease of exposition, appropriate SI interpretations for each condition are added here in parentheses but they did not appear in the actual experiment.

TABLE 2 | Examples of visible pictures for control items.

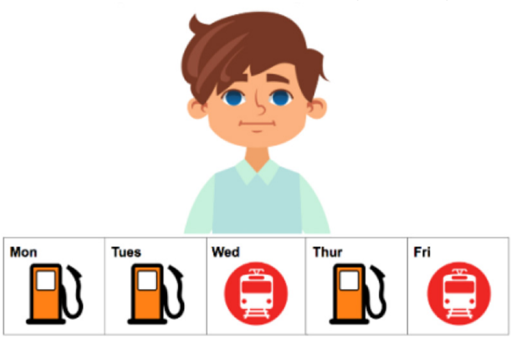
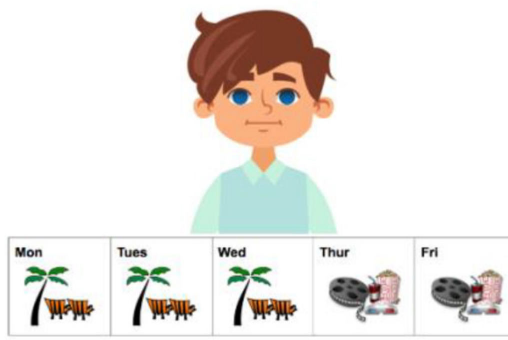
Simple affirmative sentence	Simple negated sentence
Louis went to the train station on Wednesday and Friday. 	Edward didn't go to the movies on Thursday and Friday. 

TABLE 3 | Participants' background information and proficiency scores.

	Age at testing		Years studying English		Proficiency score	
	M (SD)	**Range**	**M (SD)**	**Range**	**M (SD)**	**Range**
Native English (*n* = 26)	22.8 (5.55)	19-39	n/a	n/a	39 (1.08)	37-40
High intermediate to advanced[1] Chinese (*n* = 24)	24.2 (4.34)	18-32	14.2 (2.70)	9-18	35.04 (2.37)	30-39

[1] *The proficiency test and the categorization of advanced and intermediate learners were adopted from Cho (2017) where learners with scores above 34 were considered to be advanced learners and those who scored between 26 and 33 belonged to the intermediate level.*

on the selected picture via a mouse. A fixation cross for 1000ms was presented at the center of the screen before the display of every stimulus sentence.

Prior to the experimental trials, first, participants finished an icon recognition task which was used to make sure that participants understood the icons correctly. Secondly, participants completed six practice items using the covered-box paradigm to familiarize themselves to the task. Regarding the experimental trials, each participant finished a total of 52 items (16 targets, 16 controls, and 20 fillers) for about 15 min.

Data Analysis

For the purpose of the analysis, the percentage of selecting covered or visible pictures and response times (RTs) were the two dependent variables in the study. Responses were coded regarding whether the visible or the covered picture was selected. RTs were calculated as the time taken to select a picture. The data were trimmed in two steps. First, participants who selected pictures which obviously did not match the test sentences were planned to be removed, but this did not result in removing any data. The data were further trimmed at +/− 3 standard deviations (SDs) or more from the mean subject RTs. The trimming of extreme data points resulted in the loss of 2.6% of trials in each analysis for L1-Chinese L2-English learners and 2.4% of trials in each analysis for English speakers.

The percentage of selecting visible picture or covered box was analyzed using a generalized logistic mixed-effects regression model. The model had *Percentage* as the dependent variable, *SI type* (2 levels: DSI and ISI) and *Group* (2 levels: Native and L2) as fixed effects, participants and items as random factors.

To correct the skewed distribution of the data, RTs were log transformed and analyzed using linear mixed-effects regression model with *log-transformed RTs* as the dependent variable, *SI type* (2 levels: DSI and ISI) and *Group* (2 levels: Native and L2) as fixed effects, participants and items as random factors[8].

RESULTS[9]

Percentage of Picture Selection

To recapitulate the logic of the covered box method, if participants computed SI, they were expected to choose the visible picture when it depicted the inference and to choose the covered box when the visible picture illustrated no inference. Conversely, if the participant suspended SI, they were expected to choose the visible picture when it portrayed no inference.

When the visible picture showed an inference (as (12-13) in **Table 1**), both groups selected the visible picture 100 % of the time in the DSI condition and over 97% of the time in the ISI condition. This indicates that both native and L2 speaker groups computed DSI and ISI without any difficulties.

The percentage of selecting the covered box in the No-inference condition in DSI and ISI for both groups is visualized in **Figure 4**. When the visible picture showed an image of no-inference (as (3-4) in **Table 1**), both native and L2 groups behaved similarly by selecting the covered box about 86% of the time in the DSI condition. There was no significant difference between the two groups ($z = 0.106$, $p = 0.916$). However, the two groups differed in the ISI condition. While native speakers chose the covered box 86.2% of the time, L2 speakers selected the covered box only 72.2%, as visualized in **Figure 4**. It further suggested that in the No-inference condition of ISI, Chinese speakers were more likely to choose the visible picture than English speakers (Chinese: 27.8% vs. English: 13.8%).

Results from a generalized logistic mixed-effects model suggested a main effect of *SI type* ($\beta = 1.42$, $SE = 0.53$, $z = 2.65$, $p = 0.008$) and an interaction between *SI type* and *Group* ($\beta = -1.57$, $SE = 0.74$, $z = -2.11$, $p = 0.035$). Post-hoc comparisons indicated that the percentage of covered box selection in the No-inference condition of ISI between Chinese and English speakers was significantly different ($z = 2.082$, $p = 0.037$), as well as the percentage of Chinese speakers between DSI and ISI ($z = -2.65$, $p = 0.008$).

What stood out from the results was the higher percentage of selecting the visible picture in No-inference condition of the ISI by Chinese speakers. The visible picture in this condition represented a logical no-inference interpretation where the SI was suspended. The L2 group was significantly more likely to select the visible picture than the native speaker group in this condition. This could be interpreted in two ways. As noted in the introduction, SI suspension can be achieved through two routes: no computation of SI at all or cancelation of the SI that was initially computed. First, this could mean that L2 speakers, compared to native speakers, have more difficulties in computing ISI. Secondly, this could mean L2 speakers are better at canceling ISI. We will return to this issue in Discussion.

Response Times (RTs)

Bill et al. (2018) suggested that a comprehensive RT analysis requires the comparative examination between visible picture and covered box selection, in particular when the two types of SI are compared. The reason is that the prediction of RTs is not that RTs will be the same or different between DSI vs. ISI in that we compare two substantially different scalar items, i.e., one with negation and one without negation. Rather, the prediction is whether the overall RT patterns that are categorized by SI computation (choosing the visible picture in the Inference condition and the covered box in the No-inference condition) and SI cancelation (choosing the visible picture in the No-inference condition) are similar or different. Thus, in this study we analyze and compare RTs for selecting the visible picture and RTs for selecting the covered box. Native speakers and L2 speakers' RTs are summarized in **Tables 4**, **5**, respectively.

As shown in **Tables 4**, **5**, the mean RTs for the covered box selection in the DSI-Inference condition is 0 for both native and L2 speakers since no one selected the covered box in this condition. **Tables 4**, **5** are further visualized into **Figures 5**, **6**, respectively, by using the *ggplot2* package (Wickham, 2016).

[8]The mixed-effects generalized logistic model and linear model first included proficiency as a (continuous) fixed factor. However, the results indicated that proficiency was an insignificant factor for both models and, therefore, the simpler models without proficiency were refitted.

[9]Tables reporting fixed effects parameters appear in **Appendix A.**

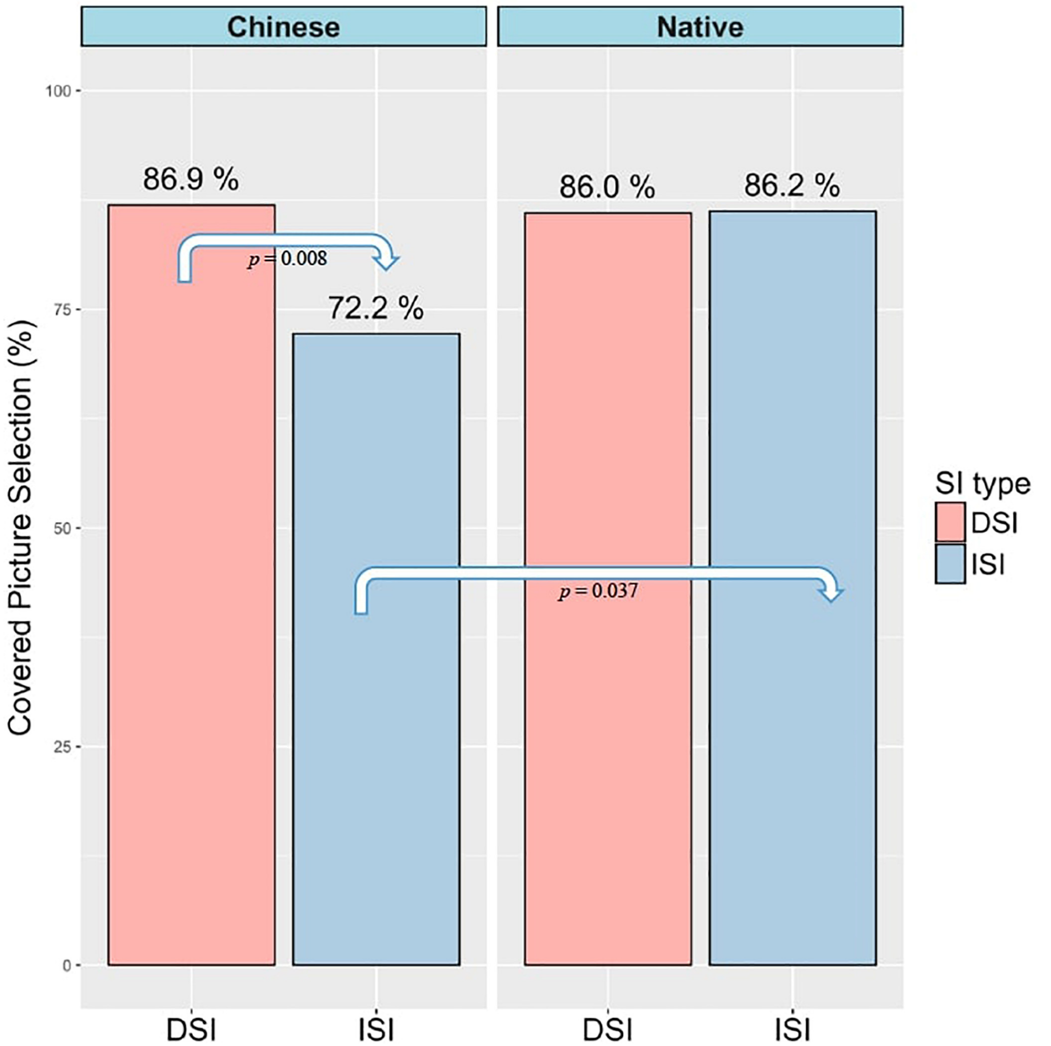


FIGURE 4 | Covered box selection percentage in the No-inference condition of DSI vs. ISI in Chinese and English groups.

Selecting the visible picture in the Inference condition of DSI was fast for both groups (English: 1273ms vs. Chinese: 1770ms). In the Inference condition of ISI, selecting the visible picture was faster than selecting the covered box for both groups. These results are not surprising since visible pictures in the Inference condition of DSI and ISI were compatible with the reading that SI inference was present. What is more interesting is the RTs in the No-inference condition (in bold in **Tables 4**, **5**) since RTs of visible picture selection represents the time to suspend SI whereas RTs of covered box selection represents the time to compute SI. It seems that both native speakers and L2 speakers were faster in selecting the covered box (computing SI) than the visible picture (suspending SI) in both DSI and ISI[10].

To investigate RTs of computing SI statistically, log-transformed RTs of the covered box selection in the No-inference condition were fitted for a linear mixed-effects regression model. Type III tests of fixed effects reported significant main effects of *SI type* ($F(1, 184.50) = 17.142$, $p < 0.001$) and *Group* ($F(1, 44.89) = 10.12$, $p = 0.002$) without significant interaction effects between the two factors. It reflected that RTs of selecting the covered box in the No-inference-visible condition of ISI were significantly longer than those of DSI ($\beta = 0.14$, $SE = 0.042$, $t = 3.288$, $p = 0.001$) and RTs of English speakers were significantly shorter than Chinese speakers ($\beta = -0.15$, $SE = 0.053$, $t = -2.863$, $p = 0.006$). It is not surprising that native speakers were faster than the L2 group. Post-hoc comparisons suggested that it took longer to select the covered box in the No-inference condition of ISI than of DSI for both groups (English: $t = -3.44$, $p < 0.001$; Chinese: $t = -3.28$, $p = 0.001$). In other words, it took longer to compute ISI than DSI for both groups when the non-dominant alternative (no-inference) meaning was explicitly offered.

[10]The outlier RTs of both groups are quite long, especially for L2 speakers whose outlier RTs are twice as long as native speakers. As a reviewer suggested, this calls for more implicit online measures (e.g., eye tracking, ERPs) since they would provide more insight on the real-time processing behavior of SI and the integration of semantic and pragmatic meanings.

TABLE 4 | Mean RTs (in ms) for selecting the visible picture vs. covered box by condition (native group).

Conditions \ Selection		Visible picture	Covered box
DSI	Inference visible picture	1273 (100%)	– (0%)
	No-inference visible picture	**2513 (14%)**	**1566 (86%)**
ISI	Inference visible picture	1810 (97.1%)	2234 (2.9%)
	No-inference visible picture	**2440 (13.8%)**	**2260 (86.2%)**

percentages of selecting visible or covered picture are in parentheses.

TABLE 5 | Mean RTs (in ms) for selecting the visible picture vs. covered box by condition (L2 group).

Conditions \ Selection		Visible picture	Covered box
DSI	Inference visible picture	1770 (100%)	– (0%)
	No-inference visible picture	**7180 (13.1%)**	**2569 (86.9%)**
ISI	Inference visible picture	2391 (97.9%)	3717 (2.1%)
	No-inference visible picture	**5084 (27.8%)**	**3499 (72.2%)**

percentages of selecting visible or covered picture are in parentheses.

Another linear mixed-effect regression model was constructed to explore RTs of suspending SI, i.e., RTs of selecting the visible picture in the No-inference condition of DSI and ISI. Type III tests of fixed effects reported significant main effects of *SI type* ($F(1, 53.474) = 5.22$, $p = 0.026$) and *Group* ($F(1, 23.916) = 14.079$, $p < 0.001$) with a marginally significant interaction between the two factors ($F(1, 55.579) = 3.439$, $p = 0.069$). It indicated that RTs of selecting the visible picture in the No-inference condition of ISI were significantly faster than that of DSI ($\beta = -0.29$, $SE = 0.09$, $t = -3.118$, $p = 0.003$). RTs of English speakers were significantly shorter than Chinese speakers ($\beta = -0.489$, $SE = 0.121$, $t = -4.031$, $p = 0.0002$). Post-hoc comparisons revealed that Chinese speakers were significantly faster in selecting the visible picture in ISI than in DSI ($t = 3.118$, $p = 0.003$) whereas English speakers' RTs did not contrast significantly between ISI and DSI ($t = 0.338$, $p = 0.736$). It means that unlike native speakers who did not show RT differences in suspending DSI vs. ISI, Chinese speakers were significantly faster to suspend ISI than DSI. The next section moves onto the discussion of these findings.

DISCUSSION

This study aimed to investigate computation and suspension of two types of SI in L2 acquisition. In this section, we discuss results of the experiment by revisiting the research questions formulated in Section 3.1.

RQ1: Do Native and L2 Speakers Differ in Generating DSI or ISI?

By employing the covered box method, the ability of generating the SI inference was indicated by participants' selection of the visible picture in the Inference condition (the inference was present in the visible picture) and the selection of the covered box in the No-inference condition (the inference was absent in the visible picture). For DSI, both groups selected the visible picture 100% when the visible picture showed the inference and preferred the covered box when the visible picture did not show the inference (English 86% and Chinese 86.9%). Moreover, RTs of selecting the visible picture in the Inference condition and the covered box in the No-inference condition revealed that the SI inference was rapidly available to both native and L2 speakers (English: visible picture 1273ms, covered box 1566 ms; Chinese: visible picture 1770 ms, covered box 2569 ms). There was no difference between the two groups in DSI computation. This seems to be in line with findings from previous studies on L2 speakers' DSI computation (Slabakova, 2010; Miller et al., 2016; Snape and Hosoi, 2018).

As for ISI, both groups selected the visible picture above 97% when the visible picture showed the inference and preferred the covered box when the visible picture did not show the inference (English: 86.2% and Chinese: 72.2%). It is interesting that English speakers were more likely to select the covered box than Chinese speakers when the visible picture showed no-inference and this difference was statistically significant ($z = 2.082$, $p = 0.037$). This seems to suggest that compared to native speakers, it is difficult for L2 speakers to compute ISI when the alternative meaning (no-inference reading in this case) is explicitly offered. In terms of response times, and similar to DSI outcomes, both groups quickly gained access to the SI inference in the Inference condition (English 1810 ms vs. Chinese 2391 ms) and the No-inference condition (English: 2260 ms vs. Chinese: 3499 ms). In short, while L2 speakers computed DSI at nativelike levels, they did not compute ISI as frequently as native speakers.

RQ2: Do Native and L2 Speakers Differ in Suspending DSI or ISI?

The ability to suspend the SI inference was suggested by the selection of the visible picture in the No-inference condition where the visible picture showed a No-inference reading.

For DSI, both groups selected the visible picture in the No-inference condition at a similar percentage (English 14% and Chinese 13.1%). However, it took significantly longer for Chinese speakers to select the visible picture in the No-inference condition (Chinese 7180ms vs. English 2513ms; $t = 4.031$, $p = 0.0002$). Since the visible picture selection percentages are similar in both native and L2 groups, RT differences between the two groups seem to be a mere quantitative difference. That is, L2 speakers are simply slower than native speakers in suspending DSI.

As for ISI, the two groups differed in the selection of the visible picture. Chinese speakers selected the visible picture at 27.8% whereas English speakers selected at 13.8%. This difference was significant ($z = 2.082$, $p = 0.037$). RT analysis also showed a difference between the two groups of selecting a visible picture (Chinese 5084 ms vs. English 2440 ms; $t = 1.989$, $p = 0.054$). Unlike the quantitative RT differences in suspending DSI, the RT differences between Chinese and English speakers

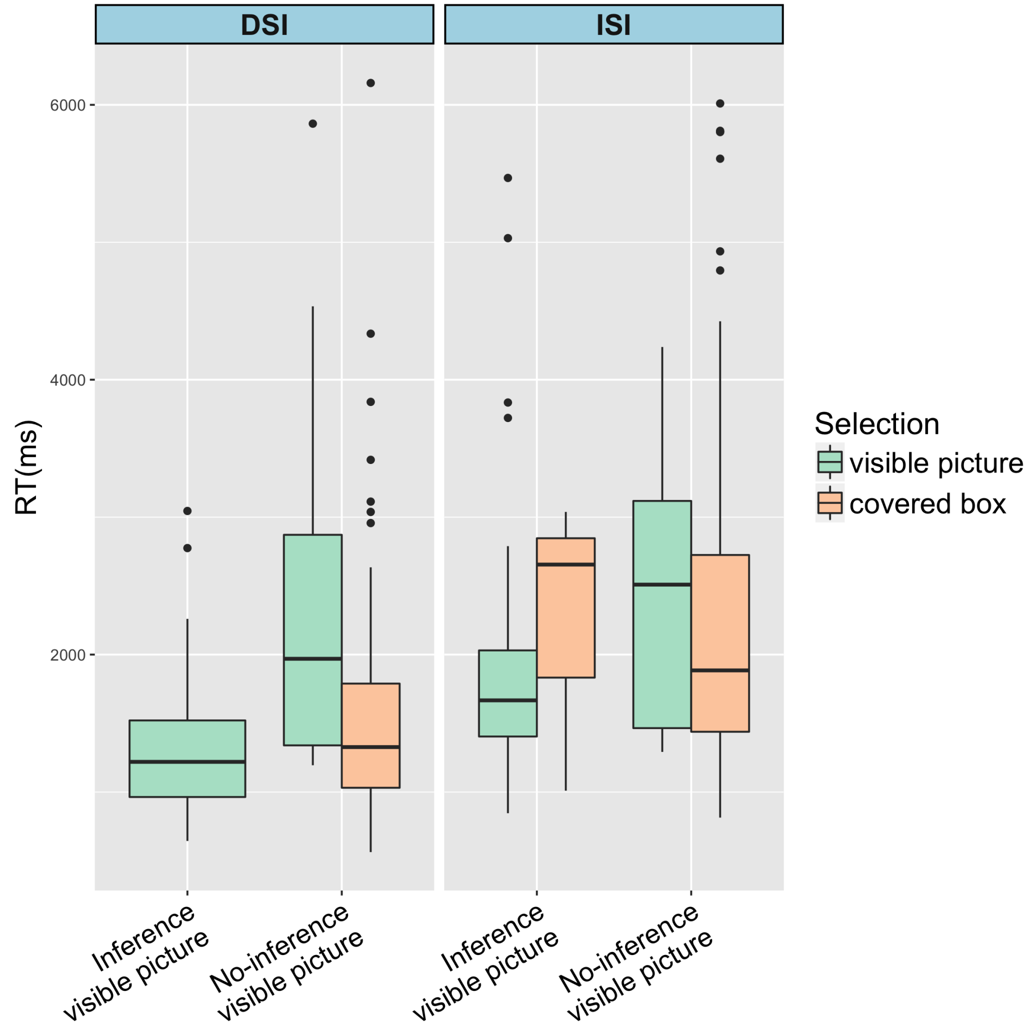


FIGURE 5 | RTs of selecting covered or visible pictures in DSI and ISI by native speakers.

in suspending ISI are qualitative, indicated by the fact that L2 speakers opted for interpretation lacking ISI more than did native speakers.

Taken together, the two types of SI inference were rapidly available to both native and L2 speakers suggested by the quick acceptance of the visible picture that was compatible with an inference reading. It also should be noted that when the visible picture displayed a No-inference reading, the rejection of the visible picture (thus selection of the covered box) was rapid as well for both groups and both types of SI. It further indicated that participants' preference of an inference reading of SI did not depend on the display of the visible picture (Inference vs. No-inference). Generating SI was overall preferred by both native and L2 speakers. The situation where we observed a significant slow-down for L2 speakers was during selection of the visible picture in the No-inference condition. In this situation, L2 speakers were faced with pressure of opposing alternatives when there was a conflict between the general preference of an inference reading and the visible No-inference reading. What is more interesting is that the pressure of the conflict seemed to be more outstanding for L2 speakers in DSI than in ISI since RTs of selecting the visible picture was significantly longer in DSI than in ISI (DSI 7180ms vs. ISI 5084ms). More importantly, since acceptance of visible pictures in the No-inference condition represents SI suspension, L2 speakers seemed to be able to 'suspend' ISI faster than DSI. Another asymmetrical behavior by L2 speakers was that L2 speakers did not compute ISI as frequently as native speakers in that L2 speakers' percentage of selecting the covered box in the No-inference condition was lower than native speakers (Chinese 72.2% vs. English 86.2%; $z = 2.082$, $p = 0.037$). According to the design of the coved box method, selecting the visible picture in the No-inference condition indicates the suspension of SI. However, as we mentioned in the introduction, there are two substantially different routes that lead to the same behavior (suspending an implicature) and we will discuss the two routes in detail in the following paragraphs. We propose that, in fact, Chinese speakers did not truly suspend the ISI inference because they did not generate the inference in the

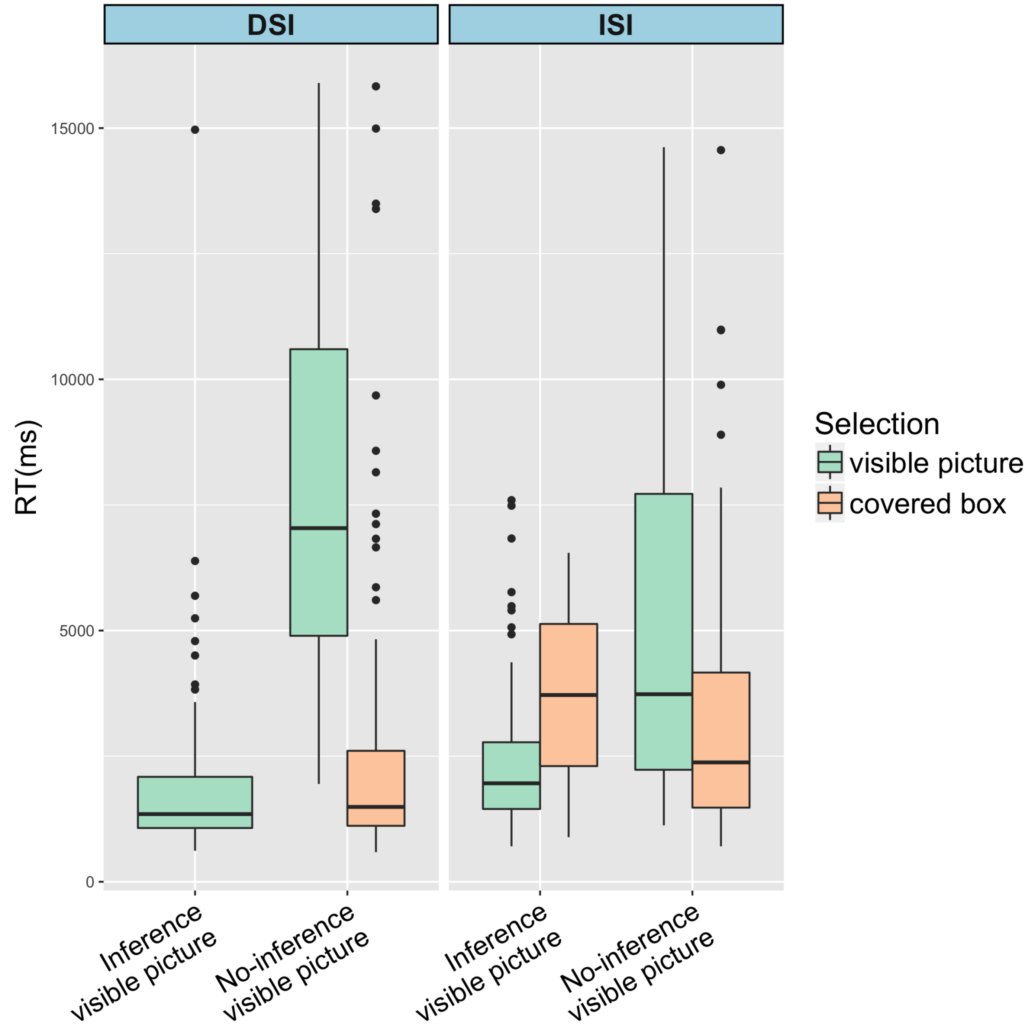


FIGURE 6 | RTs of selecting covered or visible pictures in DSI and ISI by L2 speakers.

first place, suggested by their short RTs in selecting the visible No-inference picture of ISI. Instead, they simply selected the interpretation that was visibly offered at hand (the visible No-inference picture).

In sum, comparing DSI and ISI, L2 speakers differ from native speakers in interpreting sentences containing ISI items but not DSI. While L2 speakers did not compute ISI as frequently as native speakers, they 'suspended' ISI more frequently and faster than native speakers. These asymmetries between DSI and ISI observed among L2 speakers but not among native speakers pose the following two questions. First, why does ISI computation present more challenges to L2 speakers than DSI? And secondly, why and how do L2 speakers 'suspend' ISI more frequently and faster than DSI?

As for the first question, recall the two approaches to DSI vs. ISI discussed in the introduction: the traditional view that treats DSI and ISI as the same type of implicature and the ISI as obligatory implicature (Spector, 2007; Chierchia et al., 2012). According to the traditional view, there should not be any asymmetries between DSI and ISI in their generation and suspension. While our native speaker data seem to support this view, the L2 speaker data clearly suggest that DSI and ISI do not belong to the same group of implicature. Our L2 data cannot be explained within the ISI as obligatory approach either. If DSIs are non-obligatory and ISIs are obligatory implicatures, ISIs should be computed faster and more frequently than DSIs. And DSIs should be suspended more frequently than ISIs. L2 speakers in our study showed the opposite patterns. That is, they computed DSIs more often than ISIs and suspended ISIs more often than DSIs.

To account for our results, we would like to consider differences between DSIs and ISIs in terms of the number of alternative meanings involved. Let us think about structural differences between DSI and ISI. Sentences containing a weaker term that triggers DSI as in (1a), repeated here as (16), are affirmative sentences. ISIs are triggered by negating the stronger term, as in (2a), repeated here as (17). ISIs arise in negative sentences.

(16) Bob sometimes went to school (DSI).
(17) Bob didn't always go to school (ISI).

Within alternative-based approaches to interpretation, negation is one of the linguistic phenomena where alternatives are computed in order to reach the interpretation by the hearer and numerous psycholinguistic studies have provided empirical evidence to support the claim (Fischler et al., 1983; Hasson and Glucksberg, 2006; Kaup et al., 2007; Lüdtke et al., 2008; Dale and Duran, 2011; Tian, 2014; Tian and Breheny, 2016). For example, understanding the utterance "John didn't buy a car" requires the hearer to compute the alternative, non-negated meaning "John bought a car" first and then negate it. That is, when interpreting (17), the hearer first computes the non-negated meaning "Bob always went to school" and then negates it. The negated sentence "Bob didn't always go to school" has two alternative meanings: inference ('Bob sometimes went to school.') and no-inference ('Bob possibly never went to school.'). In other words, in interpreting (17), three meanings should be computed: non-negated meaning, the literal meaning of the negated sentence, and the inferred meaning of the negated sentence. The affirmative utterance in (16), on the other hand, evokes only two alternatives: "Bob sometimes went to school" and "Bob possibly always went to school". Under the assumption that the more alternative meanings are involved in understanding an utterance, the more cognitive effort is required, (17) containing an ISI item should be more difficult to process than (16) containing a DSI item. This could explain why L2 speakers generated ISI less frequently than DSI.

This issue relates to the second question about SI suspension. As discussed in Bill et al. (2015), speakers go through the following steps in interpreting sentences containing scalar items: (1) accessing the no-inference or literal interpretation; (2) generating SI by default; (3) suspending or canceling SI if needed. We briefly mentioned in the Introduction that achieving no-inference interpretation can be done through two ways: (1) not generating SI at all; (2) canceling SI that was previously generated. Given the three steps proposed in Bill et al. (2015), it suggests that speakers who suspend SI via the first way (not generating SI at all) stop at the first step and therefore, they rapidly generate the no-inference interpretation. On the other hand, speakers who suspend SI via the second way (canceling previously computed SI) must, first, have gone through the derivation of the SI inference and then suspend it. Thus, the re-calculation of meaning is cognitively costly and thus takes longer to undergo all the steps. The asymmetry of suspending DSI and ISI observed among L2 speakers in the present study is that it took Chinese speakers significantly longer to suspend DSI than ISI (DSI 7180ms vs. ISI 5084ms; $t = 3.118$, $p = 0.003$). The longer RTs of suspending DSI by Chinese speakers suggested that they were likely to suspend DSI through the second route, i.e., generating SI and then canceling it. In other words, in reading DSI sentences presented with the alternative no-inference meaning in the visible picture, L2 speakers were able to compute the inference and the alternative reading, and then cancel the inference. Shorter RTs of ISI cancelation were due to the suspension through the first route, i.e., not generating SI at all. When the no-inference reading was offered through the visible picture in the ISI condition, it was difficult for L2 speakers to compute all the alternative readings relevant to the target sentence. So, rather than computing alternatives, L2 speakers opted for the interpretation that was visibly offered.

Finally, it is important to bear in mind that our study only examined frequency adverb scalar items; thus, our results may not be generalizable to all DSIs and ISIs. According to Van Tiel et al.'s (2016) proposal on 'scalar diversity', not all DSI items behave the same. For example, Van Tiel et al. (2016) tested 50 participants (20 males and 30 females aged 18-67) on a sentence evaluation task using Mechanical Turk. Participants saw a sentence like John says: *She is intelligent* and were asked a question like *Would you conclude from this that, according to John, she is not brilliant*? Results showed that 100 % of the participants derived SI for < cheap, free > and < sometimes, always > (i.e., 50 out of all 50 participants), while only 6 % of the participants (i.e., three out of 50 participants) computed SI for < intelligent, brilliant > (See Van Tiel et al., 2016 and Van Tiel and Schaeken, 2017 for a detailed discussion on factors influencing the rate of SI derivation).

Furthermore, we would like to consider experimental task effects and potential individual differences in interpreting data. Studies on children showed that children's logical vs. pragmatic responses differ depending on the task type, instruction, training or experimental setting (see Huang and Snedeker, 2009 for discussion on experimental task effects on inference computation in children). The patterns observed in our study may in part be due to extraneous task effects of the covered-box paradigm related to overall cognitive processing. Task effects in pragmatic inference computation suggest inference processing is closely related to cognitive abilities. In fact, recent studies have also shown that there are individual differences in L2 speakers as well as in native speakers in their computation or suspension of inferences and identified working memory as a main factor affecting inference computation (Marty and Chemla, 2013; Marty et al., 2013). Additionally, many studies have split responders into distinct groups, e.g., pragmatic responders and logical responders or responders with high or low pragmatic abilities, since participants do not have the same threshold of informativeness (Noveck and Posada, 2003; Nieuwland et al., 2010; Tavano and Kaiser, 2010). Experimental task effects and individual differences are therefore important issues for future research on pragmatic processing.

CONCLUSION

The main goal of the current study was to examine how L2 speakers compute and suspend the two types of SI, DSI and ISI. While native speakers did not compute or suspend differently between DSI and ISI, L2 speakers showed asymmetrical behaviors to DSI and ISI. More specifically, L2 speakers computed DSI more often and faster than ISI, but suspended ISI more frequently and faster than DSI. The asymmetries of the percentage and time to suspend between DSI and ISI further revealed that L2 speakers went through different routes to suspend ISI and DSI, depending on the extent of alternative meanings involved in the suspension. DSI arises in affirmative sentences while ISI arises in negated sentences which evoke computation of more alternative meanings and re-calculation. It is cognitively more demanding to generate multiple alternative meanings, re-evaluate these meanings and eventually cancel one of them.

ETHICS STATEMENT

The protocol (#2018-0330) was approved by the ED/SBS IRB (Education and Social/Behavioral Science Institutional Review Board) at the University of Wisconsin-Madison. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

SF and JC contributed to conception and design of the experiment. SF implemented the experiment and oversaw data collection. Both authors handled analysis of the experiments, contributed to manuscript revision and approved the submitted version.

FUNDING

Support for this research was provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (Grant # MSN 215833).

REFERENCES

Bacovcin, H. A., Zehr, J., and Schwarz, F. (2016). To accommodate or to ignore? The presuppositions of *again* and *continue* across contexts. *Glossa J. Gen. Linguist.* 3:16. doi: 10.5334/gjgl.402

Bergen, L., and Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1450–1460. doi: 10.1037/a0027850

Bill, C., Romoli, J., and Schwarz, F. (2015). *Are Some Presuppositions Scalar Implicatures? Assessing Evidence from Reaction Times*. Available at: https://www.academia.edu/17390559/Are_some_presuppositions_scalar_implicatures_Assessing_evidence_from_Reaction_Times (accessed February 21, 2018).

Bill, C., Romoli, J., and Schwarz, F. (2016). Scalar implicatures versus presuppositions: the view from acquisition. *Topoi* 35, 57–71. doi: 10.1007/s11245-014-9276-1

Bill, C., Romoli, J., and Schwarz, F. (2018). Processing presuppositions and implicatures: similarities and differences. *Front. Commun.* 3:44. doi: 10.3389/fcomm.2018.00044

Bott, L., Bailey, T. M., and Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *J. Mem. Lang.* 66, 123–142. doi: 10.1016/j.jml.2011.09.005

Bott, L., and Noveck, I. (2004). Some utterances are underinformative. *J. Mem. Lang.* 51, 437–457. doi: 10.1016/j.jml.2004.05.006

Breheny, R., Katsos, N., and Williams, J. (2006). Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100, 434–463. doi: 10.1016/j.cognition.2005.07.003

Carston, R. (2004). "Relevance Theory and the Saying/Implicating Distinction," in *The Handbook of Pragmatics*, eds L. Horn and G. Ward (Oxford: Blackwell), 151–181.

Chemla, E. (2009). Presuppositions of quantified sentences: experimental data. *Nat. Lang. Semantics* 17, 299–340. doi: 10.1007/s11050-009-9043-9

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., and Meroni, L. (2001). "The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures," in *Proceedings from the Annual Boston University Conference on Language Development*, Vol. 25, Boston, MA, 157–168.

Chierchia, G., Fox, D., and Spector, B. (2012). "The grammatical view of scalar implicatures and the relationship between semantics and pragmatics," in *Semantics: An International Handbook of Natural Language Meaning*, eds C. Maienborn, K. von Heusinger, and P. Portner (Berlin: Mouton de Gruyter), 2297–2331.

Cho, J. (2017). The acquisition of different types of definite noun phrases in L2-English. *Int. J. Biling.* 21, 367–382. doi: 10.1177/1367006916629577

Cremers, A., and Chemla, E. (2014). "Direct and indirect scalar implicatures share the same processing signature," in *Pragmatics, Semantics and the Case of Scalar Implicatures Palgrave Studies in Pragmatics, Language and Cognition*, ed. S. P. Reda (London: Palgrave Macmillan).201–240.

Dale, R., and Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cogn. Sci.* 35, 983–996. doi: 10.1111/j.1551-6709.2010.01164.x

Degen, J., and Tanenhaus, M. K. (2011). "Making inferences: the case of scalar implicature processing," in *Proceedings of the 33rd annual conference of the Cognitive Science Society*, Boston, MA, 3299–3304.

Dijkstra, T. (2005). "Bilingual visual word recognition and lexical access," in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. Kroll and A. de Groot (Oxford: Oxford University Press), 179–201.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., and Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology* 20, 400–409. doi: 10.1111/j.1469-8986.1983.tb00920.x

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition and Logical Form*. New York, NY: Academic Press.

Green, D. W. (1986). Control, activation and resource. *Brain Lang.* 27, 210–223. doi: 10.1016/0093-934X(86)90016-7

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Biling. Lang. Cogn.* 1, 67–81. doi: 10.1017/S1366728998000133

Grice, P. (1975). "Logic and conversation," in *The Logic of Grammar*, eds D. Davidson and G. Harman (Encino, CA: Dickenson), 64–75.

Grodner, D. J., Klein, N. M., Carbary, K. M., and Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: evidence for immediate pragmatic enrichment. *Cognition* 116, 42–55. doi: 10.1016/j.cognition.2010.03.014

Hasson, U., and Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *J. Pragmat.* 38, 1015–1032. doi: 10.1016/j.pragma.2005.12.005

Huang, Y. T., and Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: evidence from real-time spoken language comprehension. *Dev. Psychol.* 45, 1723–1739. doi: 10.1037/a0016704

Huang, Y. T., and Snedeker, J. (2011). *Logic and conversation* revisited: evidence for a division between semantic and pragmatic content in real-time language comprehension. *Lang. Cogn. Process.* 26, 1161–1172. doi: 10.1080/01690965.2010.508641

Huang, Y. T., Spelke, E., and Snedeker, J. (2013). What exactly do numbers mean? *Lang. Learn. Dev.* 9, 105–129. doi: 10.1080/15475441.2012.658731

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R., and Lüdtke, J. (2007). Experiential simulations of negated text information. *Q. J. Exp. Psychol.* 60, 976–990. doi: 10.1080/17470210600823512

Kroll, J., Bobb, S., and Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: arguments against a fixed locus of language selection in bilingual speech. *Biling. Lang. Cogn.* 9, 119–135. doi: 10.1017/S1366728906002483

Kroll, J., Dussias, P., Bice, K., and Perrotti, L. (2015). Bilingualism, mind, and brain. *Ann. Rev. Linguist.* 1, 377–394. doi: 10.1146/annurev-linguist-030514-124937

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511813313

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/5526.001.0001

Lewis, S., and Phillips, C. (2011). "Computing scalar implicatures is cost-free in supportive contexts," in *Poster Presented at the 17th Annual Conference on Architectures and Mechanisms for Language Processing*, Paris.

Lieberman, M. (2009). L2 Acquisition of scalar implicatures: Interpretation at the syntax/pragmatics interface. *Paper presented at the 10th Generative Approaches to Second Language Acquisition (GASLA) conference, University of Illinois at Urbana-Champaign*, Illinois.

Lin, Y. (2016). Processing of scalar inferences by mandarin learners of english: an online measure. *PLoS One* 11:e0145494. doi: 10.1371/journal.pone.0145494

Lüdtke, J., Friedrich, C. K., De Filippis, M., and Kaup, B. (2008). Event-related potential correlates of negation in a sentence-picture verification paradigm. *J. Cogn. Neurosci.* 20, 1355–1370. doi: 10.1162/jocn.2008.20093

Marian, V., and Spivey, M. J. (2003). Competing activation in bilingual language processing: within and between-language competition. *Biling. Lang. Cogn.* 6, 97–115. doi: 10.1017/S1366728903001068

Marty, P., and Chemla, E. (2013). Scalar implicatures: working memory and a comparison with only. *Front. Psychol.* 4:403. doi: 10.3389/fpsyg.2013.00403

Marty, P., Chemla, E., and Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua* 133, 152–163. doi: 10.1016/j.lingua.2013.03.006

Miller, D., Giancaspro, D., Iverson, M., Rothman, J., and Slabakova, R. (2016). "Not just *algunos,* but indeed *unos* L2ers can acquire scalar implicatures in L2 Spanish," in *Language Acquisition Beyond Parameters: Studies in honor of Juana M. Liceras. Studies in Bilingualism*, eds A. A. de la Fuente, E. Valenzuela, and C. Martinez Sanz (Amsterdam: John Benjamins), 125–145.

Nieuwland, M., Ditman, T., and Kuperberg, G. (2010). On the incrementality of pragmatic processing: an ERP investigation of informativeness and pragmatic abilities. *J. Mem. Lang.* 63, 324–346. doi: 10.1016/j.jml.2010.06.005

Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition* 78, 165–188. doi: 10.1016/S0010-0277(00)00114-1

Noveck, I., and Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain Lang.* 85, 203–210. doi: 10.1016/S0093-934X(03)00053-1

Papafragou, A., and Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition* 86, 253–282. doi: 10.1016/S0010-0277(02)00179-8

Politzer-Ahles, S., Fiorentino, R., Jiang, X., and Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: an event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Res.* 1490, 134–152. doi: 10.1016/j.brainres.2012.10.042

Romoli, J. (2012). *Soft but Strong: Neg-raising, Soft Triggers, and Exhaustification*. Ph. D. thesis, Harvard University, Harvard.

Romoli, J., and Schwarz, F. (2015). "An experimental comparison between presupposition and indirect scalar implicatures," in *Experimental perspectives on presupppositions: Studies in Theoretical Psycholinguistics*, ed. F. Schwarz (Cham: Springer), 215–240.

Rooth, M. (1985). *Association with Focus*. Ph. D. thesis, University of Massachusetts at Amherst, Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semantics* 1, 75–117. doi: 10.1007/BF02342617

Rooth, M. (2016). "Alternative semantics," in *The Oxford Handbook of Information Structure*, eds C. Féry and S. Ishihara (Oxford: Oxford University Press).

Schwarz, F. (2014). Presuppositions are fast, whether hard of soft- evidence from the visual world. *Proc. Salt* 24, 1–22.

Slabakova, R. (2010). Scalar implicatures in second language acquisition. *Lingua* 120, 2444–2462. doi: 10.1016/j.lingua.2009.06.005

Smith, C. L. (1980). Quantifiers and question answering in young children. *J. Exp. Child Psychol.* 30, 191–205. doi: 10.1016/0022-0965(80)90057-0

Snape, N., and Hosoi, H. (2018). Acquisition of scalar implicatures: evidence from adult Japanese L2 learners of English. *Linguist. Approaches Biling.* 8, 163–192. doi: 10.1075/lab.18010.sna

Spector, B. (2007). "Aspects of the pragmatics of plural morphology: On higher order implicatures," in *Presupposition and Implicature in Compositional Semantics*, eds U. Sauerland and P. Stateva (London: Palgrave Macmillan), 243–281. doi: 10.1057/9780230210752_9

Sperber, D., and Wilson, D. (1986/1995). *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

Tavano, E., and Kaiser, E. (2010). "Processing scalar implicature: what can individual differences tell us?" in *Proceedings of the 33rd Annual Penn Linguistics Colloquium*, Vol. 16, California, 214–224.

Tian, Y. (2014). *Negation Processing: A Dynamic Pragmatic Account*. Ph. D. thesis, University College London, London.

Tian, Y., and Breheny, R. (2016). "Dynamic Pragmatic View of Negation Processing," in *Negation and Polarity: Experimental Perspectives Language, Cognition, and Mind*, Vol. 1, eds P. Larrivée and C. Lee (Cham: Springer), 21–43. doi: 10.1007/978-3-319-17464-8_2

Van Tiel, B., and Schaeken, W. (2017). Processing conversational implicatures: alternatives and counterfactual reasoning. *Cogn. Sci.* 41, 1119–1154. doi: 10.1111/cogs.12362

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., and Geurts, B. (2016). Scalar diversity. *J. Semantics* 33, 137–175.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-3-319-24277-4

Zehr, J., Bill, C., Tieu, L., Romoli, J., and Schwarz, F. (2016). Presupposition projection from the scope of none: universal, existential, or both? *Proc. Salt* 26, 754–774. doi: 10.3765/salt.v26i0.3837

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX A

Fixed Effects Parameters for Generalized Logistic Mixed-Effects Model and Linear Mixed-Effects Model

TABLE 6 | Fixed effects Estimates and Standard Errors for generalized logistic mixed-effects model of percentage of selecting covered or visible pictures.

	Estimate	Std. Error	*z* value	*p* value
SI type (DSI)	1.42	0.53	2.650	0.008**
Group (English)	−0.09	0.81	−0.11	0.916
SI type : Group	−1.57	0.74	−2.11	0.035*

TABLE 7 | Fixed effects Estimates and Standard Errors for linear mixed-effects model of selecting the covered box in the No-inference condition for DSI & ISI.

	Estimate	Std. Error	*Df*	*t* value	*p* value
SI type (DSI)	0.14	0.042	258.2	3.288	0.001**
Group (English)	−0.15	0.053	64.0	−2.863	0.006**
SI type : Group	−0.004	0.048	255.2	−0.103	0.918

TABLE 8 | Fixed effects Estimates and Standard Errors for linear mixed-effects model of selecting the visible picture in the No-inference condition for DSI & ISI.

	Estimate	Std. Error	*Df*	*t* value	*p* value
SI type (DSI)	−0.29	0.09	58.01	−3.118	0.003**
Group (English)	−0.489	0.121	43.09	−4.031	0.0002***
SI type : Group	0.259	0.139	55.58	1.854	0.069

frontiers in Communication

ORIGINAL RESEARCH
published: 24 April 2019
doi: 10.3389/fcomm.2019.00015

Presupposition Accommodation of the German Additive Particle *auch* (= "too")

Mira Grubic and Marta Wierzba*

Linguistics Department, University of Potsdam, Potsdam, Germany

Presupposition triggers differ with respect to whether their presupposition is easily accommodatable. The presupposition of focus-sensitive additive particles like *also* or *too* is often classified as hard to accommodate, i.e., these triggers are infelicitous if their presupposition is not entailed by the immediate linguistic or non-linguistic context. We tested two competing accounts for the German additive particle *auch* concerning this requirement: First, that it requires a focus alternative to the whole proposition to be salient, and second, that it merely requires an alternative to the focused constituent (e.g., an individual) to be salient. We conducted two experiments involving felicity judgments as well as questions asking for the truth of the presupposition to be accommodated. Our results suggest that the latter account is too weak: mere previous mention of a potential alternative to the focused constituent is not enough to license the use of *auch*. However, our results also suggest that the former account is too strong: when an alternative of the focused constituent is prementioned and certain other accommodation-enhancing factors are present, the context does not have to entail the presupposed proposition. We tested the following two potentially accommodation-enhancing factors: First, whether the discourse can be construed to be from the perspective of the individual that the presupposition is about, and second, whether the presupposition is needed to establish coherence between the host sentence of the additive particle and the preceding context. The factor coherence was found to play a significant role. Our results thus corroborate the results of other researchers showing that discourse participants go to great lengths in order to identify a potential presupposition to accommodate, and we contribute to these results by showing that coherence is one of the factors that enhance accommodation.

Keywords: alternatives, additive particles, presupposition, anaphoricity, accommodation, experimental data, German

Edited by:
Sophie Repp,
Universität zu Köln, Germany

Reviewed by:
Nadine Bade,
University of Tübingen, Germany
Florian Schwarz,
University of Pennsylvania,
United States

****Correspondence:***
Mira Grubic
grubic@uni-potsdam.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *30 November 2018*
Accepted: *02 April 2019*
Published: *24 April 2019*

Citation:
Grubic M and Wierzba M (2019) Presupposition Accommodation of the German Additive Particle auch (= "too"). Front. Commun. 4:15. doi: 10.3389/fcomm.2019.00015

1. INTRODUCTION

Additive particles belong to the class of alternative-sensitive particles, i.e., they interact with alternatives. For example, the presupposition contributed by *also* in (1) changes depending on the location of focus (Rooth, 1996, p. 272), which is marked by prosodic prominence in English (indicated by small caps): In (1a), *also* contributes the presupposition that John introduced somebody else to Sue, whereas in (1b) John introduced Bill to somebody else.

(1) a. John also introduced BILL to Sue.
b. John also introduced Bill to SUE.

In an Alternative Semantics account, focus is assumed to indicate salient alternatives (Rooth, 1985, 1992, 1996, i.a.). Apart from their ordinary semantic value, all expressions also have a focus value (indicated by the superscript f in (2ac)). The focus value of the focused constituent is a set of alternatives of the same type as the denotation of the focused constituent (2a). For out-of-focus constituents, the focus value is a set containing only the denotation of the constituent itself (2b). Focus values can be composed via pointwise functional application to derive the focus value of the whole proposition (2c), a set of propositions which differ only in that the focused constituent is replaced by its alternatives.

(2) John introduced Bill to SUE.
a. $[\![\text{SUE}]\!]^f$ = {Sue, Mary, Henry, ...}
b. $[\![\text{Bill}]\!]^f$ = {Bill}
c. $[\![\text{John introduced Bill to SUE}]\!]^f$ = {John introduced Bill to Sue, John introduced Bill to Mary, John introduced Bill to Henry, ...}

Kripke (2009) proposed that additive particles as in (1) do not merely contribute an existential presupposition ("one of the propositional focus alternatives is true"), but give rise to an anaphoric presupposition amounting to the requirement that one of the propositional focus alternatives is prementioned. For example, in (3), the first sentence entails that John is afraid of something/somebody. This information is not enough to license the use of *also*, which seems to require that the context specifies who exactly the other individual that John is afraid of is.

(3) John is a coward. He is even/# also afraid of PETER!

In the literature, two main accounts of the anaphoricity of additive particles are prevalent: According to the first account, a propositional alternative has to be prementioned or entailed by the immediately preceding context (Asher and Lascarides, 1998; Beaver and Zeevat, 2007; Roberts, 2010; Tonhauser et al., 2013, i.a). We will call this account the SALIENT PROPOSITION account. A formal proposal of this can be found in (4) (from Chemla and Schlenker, 2012, p. 190): *too* requires that a proposition is salient which is (i) true in the evaluation world w, and which (ii) contextually entails one of the focus alternatives of the proposition which is logically independent of the proposition itself[1].

(4) $[\![\text{too}_i \text{ IP}]\!]_0^{s,w}$ = # unless
(i) s(i) denotes a proposition which is true at w, i.e., s(i)(w) = true; and
(ii) for some proposition α in $[\![\text{IP}]\!]_f^{s,w}$,
(a) α is [logically independent] from $[\![\text{IP}]\!]_0^{s,w}$, and
(b) relative to the context set, s(i) entails α
If $[\![\text{too}_i \text{ IP}]\!]_0^{s,w} \neq \#$, then $[\![\text{too}_i \text{ IP}]\!]_0^{s,w} = [\![\text{IP}]\!]_0^{s,w}$

An account making the same predictions but for different reasons is that of Ruys (2015): He proposes that additive particles in fact merely contribute an existential presupposition that one of the other focus alternatives needs to be true (5). The anaphoricity is due to the fact that *too* is focus-sensitive and that focus is anaphoric (though Ruys formalizes this in terms of Givenness; see Schwarzschild 1999). The combination of focus anaphoricity and existential presupposition explains why *too* is licit in a subset of possible focus contexts, namely only those in which another alternative is previously entailed to be true.

(5) $\varphi[\alpha_F]$ too = $\varphi[\alpha_F]$,
defined iff $\exists x[x \neq \alpha \;\&\; \varphi[x]]$

The second main account of anaphoricity is one which suggests that additive particles merely require the salience of a focus alternative of the focused constituent (Corblin, 1991; Heim, 1992; Geurts and van der Sandt, 2004). Since in most examples, this is an individual, we will call this the SALIENT INDIVIDUAL account. (6) is the variant of this account proposed by Heim (1992). In this case, the presupposition involves an individual pronoun x_i which requires an antecedent in the immediately preceding context. This individual pronoun is responsible for the anaphoricity of *too* and other additive particles.

(6) $\phi[\alpha_F]$ too_i = $\phi[\alpha_F]$,
defined iff $x_i \neq \alpha \;\&\; \phi[x_i]$

A further variant of this account is that of Geurts and van der Sandt (2004), who propose that sentence (7) actually has two presuppositions, (7a) and (7b)[2] (Geurts and van der Sandt, 2004, p. 31).

(7) The VICAR is depressed, too.
a. There is some person x other than the vicar
b. x is depressed

They propose that the first presupposition is hard to accommodate because it is poor in descriptive content, whereas the second presupposition can be accommodated. The underlying assumption is that the reason why some presuppositions are hard to accommodate whereas others are easy to accommodate is that the former lack a rich descriptive content—the hearer thus doesn't have enough information to know what exactly to accommodate (van der Sandt, 1992).

We test these two accounts of the anaphoricity of additive particles in two experimental studies using semantic judgments and questions targeting the presupposition, following up on a previous study in Grubic (forthcoming). This paper is structured as follows. First, we discuss experimental work on presupposition

[1]Two propositions are logically independent if neither entails the other. This is needed to rule out sentences as in (i).

(i) a. John and Bill came to the party, and JOHN came too.
b. John came to the party, and JOHN AND BILL came too.

[2]They attribute the latter presupposition to focus, via their background presupposition rule (BPR). Note however that the BPR is highly disputed (see the replies in the same volume, e.g., Büring 2004, p. 72; Jäger 2004, p. 112).

accommodation in section 2. Thereby, section 2.1 introduces the topic and shows that the German triggers discussed here behave the same as their English counterparts. Section 2.2 discusses experimental work on the accommodation of the presupposition of additive particles. These studies find that participants are willing to reinterpret the test sentences or accommodate that something in the context is the required antecedent, in order to assure that the presupposition is satisfied in the context. There are only few studies which would allow us to decide between the accounts mentioned above, and they do not investigate which further factors play a role for accommodation. Section 2.3 discusses a previous experiment by the first author on the presupposition of German *auch* (= "too"). The results seemed, on first glance, to confirm the predictions of the SALIENT INDIVIDUAL account, but additive particles in neutral contexts (i.e., where neither of the presuppositions proposed by Geurts and van der Sandt for additive particles are satisfied) received unexpectedly high felicity ratings. Section 3 reports two new follow-up experiments that we conducted. The first experiment, discussed in section 3.1, improves the methodology and stimuli, and controls for two different kinds of possible confounds: (i) whether the presupposition of *auch*, when accommodated, enhances coherence, and (ii) whether the context is from the point of view of the individual that the presupposition is about. The results show that at least one of these factors, namely coherence, facilitates accommodation. However, the experiment retained the problem that *auch* was judged to be quite acceptable in neutral contexts. For this reason, two changes were made to the materials in the second experiment, discussed in section 3.2, to see whether this has any influence on the felicity of *auch* in neutral contexts. First, to see whether participants correctly identify the associate of *auch*, a third type of test sentence was added in which *auch* unambiguously associates with the subject. A similar pattern was found for the unambiguous version, showing that this was not the problem. Second, items in which the presupposition was something very frequent or likely were changed, in order to see whether likelihood of the presupposition plays a role. The result patterns in the second experiment were indeed clearer, arguably due to the elimination of this potential confound. Again we find an effect of coherence, and find that in contexts without any potentially accommodation-enhancing factors, the results look as predicted by the SALIENT PROPOSITION account.

2. BACKGROUND

2.1. Presupposition Accommodation

Presupposition triggers differ with respect to the ease with which their presupposition can be accommodated, i.e., how felicitous the utterance is when the presupposition is new to the addressee. For example, the presupposition of the possessive construction is easy to accommodate, while those of additive particles and pronouns like *he* are not easily accommodatable (e.g., Beaver and Zeevat, 2007). Since the German equivalents of these three triggers will play a role in our experiments, this is demonstrated here for them by checking whether the respective trigger is infelicitous in a *neutral* context, a context that doesn't entail the presupposition, in contrast to a minimally different *positive* context which does entail the presupposition (test and examples taken from Tonhauser et al., 2013). These sentences were tested as part of our fillers in both experiments, and, as expected, the sentences (a) in which the context does not entail the presupposition were rated felicitous in (8) (with a mean rating of 4.12 on a 5-point scale[3]), but substantially worse in (9)–(10) (mean rating: 2.16/2.79, respectively). All sentences were judged to be felicitous in the minimally different contexts in (b) entailing the presupposition (with mean ratings of 4.74/4.00/4.82, respectively, on a 5-point scale).

(8) *Possessive*: Annika has to conduct job interviews but is behind schedule. The current candidate is in a hurry, she says: "I have to go ..."

a. Ich muss **meine Tochter** von der Kita abholen.
I have.to my daughter from the kindergarden pick.up
"I have to pick up my daughter from kindergarden."

b. Ich habe eine Tochter, und ich muss **meine Tochter** von der Kita abholen.
I have a daughter and I have.to my daughter from the kindergarden pick.up
"I have a daughter, and I have to pick up my daughter from kindergarden."

(9) *Additive*: Mia is eating a salad on the bus. A stranger sits down next to her and says:

a. Unser Busfahrer isst **auch** ein Brötchen.
our bus.driver eats also a bun
"Our bus driver is eating a bun, too."

b. Unser Busfahrer isst **auch** einen Salat.
our bus.driver eats also a salad
"Our bus driver is eating a salad, too."

(10) *3rd person pronoun*: Marko has to give a presentation about his family in school. He starts as follows:

a. **Er** ist Bauer.
he is farmer
"He is a farmer."

b. Mein Vater heißt Hans. **Er** ist Bauer.
my father is.named Hans he is farmer
"My father's name is Hans. He is a farmer."

In the following, the previous experimental literature on additive particles and presupposition accommodation is discussed, before introducing our own experiments, in which we compare additive particles to possessives with a third person pronoun possessor.

2.2. Accommodation and Additive Particles

Often, experiments on the accommodatability of the presupposition of additive particles have focused on whether some other linguistic material provided in the discourse can be interpreted as the required antecedent. For example,

[3]The mean ratings reported here are from Experiment 1, those of Experiment 2 were very similar.

Singh et al. (2016) test sentences like the four possible variants of (11). They propose that if the presupposition of *too* (here "John went swimming") is made very plausible by the assertion that he went to the pool, accommodation is possible (acceptability was 94%), whereas when it is made implausible by the assertion that he went to the mall, the sentences are less acceptable (40%)[4].

(11) John will go to {the pool/the mall} this morning. Peter will go swimming {tomorrow/too} after he gets back from school.

Schwarz (2007) investigates whether in German sentences with ambiguous relative clauses (subject vs. object relative clause readings), the syntactically dispreferred object relative clause reading is more readily available when it helps make sense of an additive particle. For example, the additive particle in (12) requires an antecedent entailing that somebody else saw the woman. This is satisfied in (12) under its object relative clause reading ("The woman that the girl saw had been seen by the man, too"), but not under its subject relative clause reading ("The woman that saw the girl had been seen by the man, too"). He finds that in order to find suitable antecedents for *auch* (= "also/too"), German native speakers can override syntactic processing preferences.

(12) Die Frau, die das Mädchen sah, hatte
the woman$_{N/A}$ who$_{N/A}$ the girl$_{N/A}$ saw had
{auch/vorher} der Mann gesehen.
also/before the man$_{N}$ seen
"The woman that {saw the girl/the girl saw} had been seen by the man {too/before}."

Chemla and Schlenker (2012) investigate whether in sentences like (13), participants conclude that it is reasonable for Anne to study abroad according to the speaker, i.e., that the presupposition of *aussi* ("Anne makes a reasonable decision") is contextually entailed by the overt antecedent ("Anne decides to study abroad"). This was tested for a variety of different complex sentences, including ones where the "antecedent" followed the host sentence of *aussi*.

(13) Si Anne décide de faire ses études à
if Anne decides to do her studies at
l'étranger, son frère va lui aussi prendre
the-foreign.country her brother will he also take
une décision raisonnable.
a decision reasonable
"If Anne decides to study abroad, her brother too will make a reasonable decision."

The results of these experiments suggest that hearers are willing to go to great lengths to find a suitable antecedent in the context. In particular, they suggest that hearers can, if it is plausible enough, accommodate that some preceding linguistic material is in fact the required antecedent. These experiments do not, however, test whether the presupposed proposition itself can be accommodated.

Out-of-the-blue, or with a preceding so-called *neutral* context—one that does not entail the presupposition—sentences with additive particles have been found to be degraded. For example, Tiemann (2014) tests sentences like (15) in the three kinds of contexts in (14). She reports a mean acceptability rating of 1.7 on a 4 point scale (1 = unacceptable, 4 = fully acceptable) for German *auch* in neutral contexts (Tiemann 2014, p. 76; see also Tiemann et al., 2011). The test sentences were judged significantly worse in negative contexts (mean acceptability: 1.3), and significantly better in positive contexts (3.7).

(14) a. *Positive context:* Fritz is cooking soup with Tina today.
b. *Negative context:* Nobody is cooking soup with Tina today.
c. *Neutral context:* Nobody is eating soup with Tina today.

(15) Sie hofft, dass auch SUSANNE eine Suppe mit ihr
She hopes that too Susanne a soup with her
kocht und kauft dafür Zutaten.
cooks and buys for.it ingredients
"She hopes that Susanne will cook a soup with her, too, and buys ingredients for it."

These results thus corroborate the claim that additive particles require their presuppositions to be entailed by the preceding context, i.e., that a presupposed proposition cannot be accommodated.

Note however, that there are other experimental results suggesting that when a subsentential focus alternative is in the preceding context, accommodation is possible, e.g., the eye-tracking experiments reported in Kim (2012, p. 64–84, 118–131) and Kim (2015, p. 120–128). For example, in (16), in order to make sense of the additive particle, the experiment participants were required to accommodate that Jane has some pears and some oranges, an inference that is not entailed by anything in the preceding context (Kim, 2012, p. 65). Participants were asked to pick a picture displaying what Jane has, (i) only apples, (ii) pears and oranges, (iii) pears, oranges and apples, or (iv) only oranges. Importantly, the participants were provided with the option of ignoring the presupposition by picking the picture containing only apples[5]. Kim's results indicate

[4]It was however not controlled for, in Singh et al. (2016)'s experiment, whether the test sentence and its antecedent were actually sufficiently different, e.g., whether something like (i) is felicitous at all (suggesting that *going to the pool* $\neq$ *swimming*). It might be that activity mentioned in the antecedent is too similar—near synonymous—to that mentioned in the test sentence in some of the examples tested by Singh et al. (2016).

(i) John will go to the pool this morning. Peter will go swimming.

[5]Schwarz (2015, p. 92–97) and Kim (2012, p. 110–118), Kim (2015, p. 114–119) report similar eye-tracking studies in which participants clicked on pictures which suggested that they accommodate a parallel proposition based on what other previously mentioned individuals did. However, these studies did not allow for participants to click on an picture corresponding to just the asserted part, i.e., choosing to ignore the presupposition was not a possible choice. See also Romoli et al. (2015) and Schwarz (2015, p. 98–105) for similar experiments, but with overt antecedents.

that the participants accommodate the presupposition (in about 80% of cases in the experiments for which she provides the response data), and her eye-movement data suggests that the presupposition is processed immediately as soon as the additive particle is heard.

(16) [Mark has some pears and some oranges.]
Jane also has some APPLES.

A similar experiment is reported in Gotzner and Spalek (2014) for German *auch* (= "also"/"too"), compared to *nur* (= "only") and focus. The context introduced two individuals (in (17), a judge and a witness), one of whom served as the focus in the target sentence. Participants were asked to judge the felicity of the short story on a scale from 1 (not at all acceptable) to 7 (very acceptable), and to judge the truth of a sentence with respect to the story, for example for (17) "the witness believed the defendant". The type of accent on the focused constituent was manipulated, but did not yield a significant difference in the *auch* conditions.

(17) (The judge and witness followed the argument.)
Auch der RICHTER glaubte dem Angeklagten.
also the judge believed the defendant
"Also the judge believed the defendant."
(Continuation sentence: He announced the verdict.)

There was no significant difference between the felicity ratings for the different conditions. The *auch*-sentences received a mean felicity rating of 5.7. Concerning the truth value judgments, the mean percentage of TRUE responses for the *auch*-conditions was around 74%. Both results suggest that accommodation of the presupposition of *auch* is possible. The felicity judgment results additionally suggest that if there is any additional cost of accommodation, it either does not effect felicity ratings or is masked by the fact that the other conditions also involved making additional inferences of some kind (presupposition accommodation with *nur*, conversational implicatures with the bare focus).

Note however that there are also experiments which are similar to the ones by Kim (2012, 2015) and Gotzner and Spalek (2014) and nevertheless report differences in accommodation between different triggers. Domaneschi et al. (2014) report the results of an experiment on Italian in which participants heard short stories, each containing several different kinds of presupposition triggers, among them additive particles. For additives, since the stories contained other individuals, accommodation is predicted by the SALIENT INDIVIDUAL account, but not the SALIENT PROPOSITION account. The participants then had to answer questions about the story, some of which tested whether the presuppositions were accommodated. They were distracted from this task by an additional memory task. Domaneschi et al. (2014) found that whereas with definite descriptions and factive verbs, presupposed information was accommodated and recalled in 87% and 88% of cases, respectively, this was only the case in 57/59% of cases for iteratives and additive particles (*anche, pure* ("also"), and *persino* ("even"), Domaneschi, p.c.). Thus, while the accommodation rate for additives was much higher than in Tiemann's experiment, it was still lower than for other presuppositions.

The contrast between the results of Tiemann (2014)'s experiment and the ones of Kim (2012, 2015), Gotzner and Spalek (2014), and Domaneschi et al. (2014) suggest that the SALIENT INDIVIDUAL account might be on the right track, since in the latter experiments, accommodation was possible to some extent even though no salient parallel proposition was entailed by the previous context. Instead, alternatives to the focused constituent were contextually available in these experiments. In Tiemann's experiment, neither was available, and accommodation thus not possible. Note however that Domaneschi et al. still find differences between different triggers, nevertheless, and that, as Beaver and Zeevat (2007) demonstrate with their example (18), not all environments which license pronouns also license an additive particle. While after *Jane likes Bill*, it is possible to refer back to Jane using the pronoun *she*, it is hard, according to Beaver and Zeevat, to accommodate that Jane is having dinner in New York. Further factors thus seem to play a role.

(18) ?Jane likes Bill. Bill is having dinner in New York, too.

Thus, to sum up, there are up to now only few experiments testing the predictions of the SALIENT INDIVIDUAL/SALIENT PROPOSITION accounts, and none—to our knowledge—which test whether any further factors play a role. We report, in section 2.3, the result of a previous experiment in which accommodation of the additive presupposition seemed easier than expected, even in neutral contexts, and then, in the main part of this paper, we present and discuss our follow-up experiments in which we additionally manipulated the following factors: (i) whether accommodation enhances coherence, and (ii) whether or not the context is from the point of view of the individual that the presupposition is about.

2.3. Comparing Possessives and Additive Particles

In this section, we will summarize in some detail the results of an experiment by Grubic (forthcoming). Our new experiments that we will present in section 3 are designed as follow-ups to this study.

Grubic (forthcoming) compared the behavior of possessive pronouns and the additive particle *auch* in three different types of context. In the first one, called *positive context*, both potential presuppositions of the possessive/additive were explicitly satisfied by the context (1. that there is a salient individual, 2. that this individual has the relevant property). In the second type, the *neutral context*, neither of these presupposition was satisfied. In the third type of context, the *mixed context*, only one of the presuppositions (that there is a salient individual) was satisfied, but not the other (that the individual has the relevant property). Examples of the context types is given in (19). The target sentences are illustrated in (20).

(19) a. *Positive context:* Hannes met his new classmate Isa. He wears glasses, and is sometimes teased because of them. She seemed very likeable...

✓ *1. There is a salient individual (Hannes).*
✓ *2. This individual wears glasses (explicitly mentioned).*

b. *Neutral context:* Yesterday, there was a new student in class. She seemed very likeable...

✗ *1. There is a salient individual.*
✗ *2. The individual wears glasses.*

c. *Mixed context:* Hannes met his new classmate Isa. She seemed very likeable, ...

✓ *1. There is a salient individual (Hannes).*
✗ *2. This individual wears glasses.*

(20) a. *Target sentence with possessive trigger:*

...weil sie **seine** Brille gelobt hat.
because she his glasses complimented has

"...because she complimented **his** glasses."

b. *Target sentence with additive trigger:*

...weil sie **auch** eine Brille hat.
because she also DET.INDEF glasses has

"...because she wears glasses, **too**."

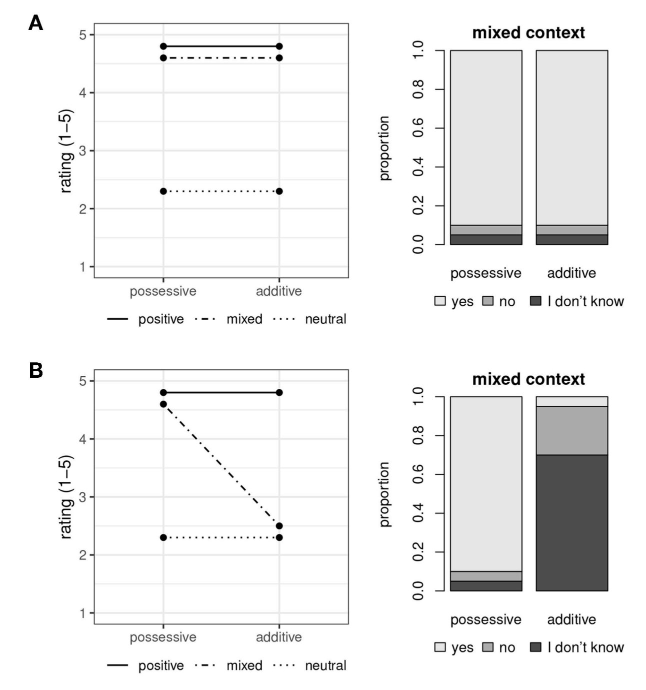


FIGURE 1 | The SALIENT INDIVIDUAL account **(A)** predicts no interaction, but a significant difference between mixed/positive and neutral context. The SALIENT PROPOSITION account **(B)** predicts an interaction between TRIGGER TYPE (possessive/additive) and CONTEXT (mixed/positive/neutral).

The participants' task was to rate the felicity of the items, and to answer a question that checked whether the second presupposition had indeed been accommodated. For example, after reading the mixed context in (19c) and the target sentence with an additive trigger in (20b), the participant was asked to judge whether the sentence makes sense in the context provided ("How good is the sentence in this context? Please rate it on a scale from 1 (not good at all) to 5 (very good)"), and to answer the question "Does Hannes wear glasses?" (Yes/No/I don't know).

Third-person possessive pronouns are particularly well suited as a baseline for the experiment because they involve presuppositions that are similar to those that the salient individual account predicts for additives. First, being third-person pronouns, they require a salient antecedent; e.g., in (20), *seine Brille* ("his glasses") presupposes that there is a salient male individual. As discussed above in section 2.1, it is an uncontroversial assumption (which is also supported by the findings in our filler items) that this presupposition cannot be accommodated. Second, being possessives, they presuppose a possessive relation between that individual and their complement NP; e.g., in (20), *seine Brille* ("his glasses") presupposes that the salient male individual has glasses. As discussed above in section 2.1, it is an uncontroversial assumption (again, supported by our filler items) that this presupposition can be accommodated very easily. Thus, possessive pronouns should receive low ratings in the neutral context, in which no suitable individual is made salient by the context. In the positive context, they should receive high ratings, because both presuppositions are satisfied by the context. In the mixed context, they should also receive high ratings, because the salient individual presupposition is satisfied, and the possessive relation can be easily accommodated.

As for additives, the predictions differ between the approaches discussed above. If it is correct that additives are similar to possessives in that the presence of a salient individual is sufficient and the relevant proposition can be accommodated, they should show the same pattern as possessive pronouns: they should also be acceptable in both the mixed and the positive context. TRIGGER TYPE (possessive vs. additive) should thus not interact with CONTEXT (mixed vs. positive/neutral). In addition, the proportion of *yes*-answers to the question asking whether the presupposition is fulfilled should be very high for both possessives and additives.

If, however, the relevant proposition needs to be explicitly expressed in the context, they should only be acceptable in the positive context. In that case, TRIGGER TYPE should interact with CONTEXT in the following direction: the felicity difference between the positive and the mixed context should be larger for additives than for possessives, and the difference between the neutral and the mixed context should be smaller for additives than for possessives. Concerning the answers to the content question, additives in the mixed context should give rise to a high proportion of "no" or "I don't know" answers, indicating that the presupposition cannot be accommodated. The predictions are illustrated in **Figure 1**.

The results are presented in **Figure 2**. The left plot shows felicity ratings on a 1–5 scale. The interaction between TRIGGER TYPE (possessive vs. additive) and CONTEXT was significant when comparing the mixed to the neutral context: the difference between the two contexts was smaller for additives than possessives. No significant interaction was found when comparing the mixed to the positive context (see Grubic forthcoming for statistical details). The right plot shows

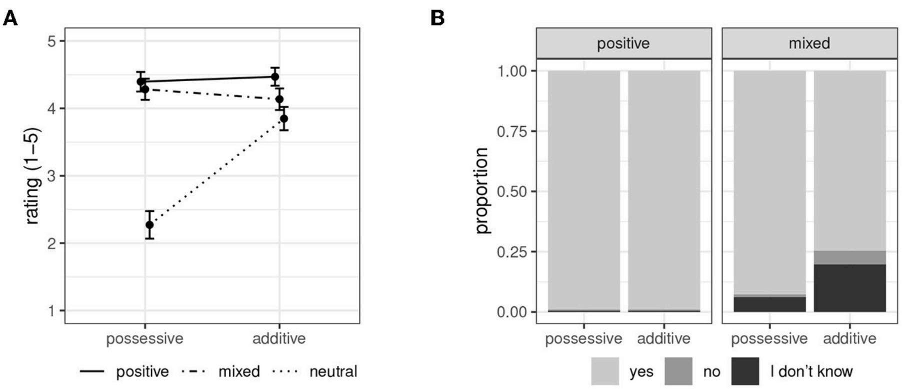


FIGURE 2 | Results of the experiment reported in Grubic (forthcoming): felicity ratings with 95% confidence intervals **(A)** and answer proportions **(B)**.

proportions of answers to the content question. They are only shown for the positive and mixed context here, in which the same question was asked ("Does Hannes wear glasses?" in (20a) and (20c))[6].

The results show that for additives, the felicity ratings in the mixed context are closer to the ratings in the neutral context than for possessives. However, this difference is mainly driven by an unexpected behavior of the additive particles in the neutral context: their ratings were much higher here than for possessives. Neither of the two approaches predicted this pattern. As for the answers to the content questions, they were almost always answered by "yes" in the positive context. In the mixed context, a higher proportion of "no" and "I don't know" answers was found, but a clear majority of trials was answered by "yes." This suggests an overall high rate of accommodation, in line with the predictions of the SALIENT INDIVIDUAL account, which was however a bit lower for additives than for possessives.

3. EXPERIMENTS

3.1. Experiment 1

3.1.1. Motivation

In Grubic (forthcoming)'s experiment discussed above, an interesting result was that one of the contexts that was intended as a negative baseline (the neutral context) did not show the expected results with the additive particle. Rather than rejecting sentences containing *auch* without a salient parallel proposition nor a salient individual in the context, participants rated them as highly acceptable.

We hypothesized that this might be due to the written modality of the experiment. Reading the target sentence with an implicit prosody that differed from the intended one could have altered the interpretation: instead of the intended association with the subject (which requires stress on the additive particle), participants might have read the sentences with stress on the object, leading to association of the additive particle with the object. In some of the items, this might result in a plausible reading in the neutral context, potentially increasing felicity.

We additionally changed some of the neutral contexts. It was noted in Grubic (forthcoming) that the neutral contexts in which the test sentences received the highest ratings often involved quantifiers, where the implicit restrictor of the quantifier might serve as an antecedent. For example, it seems relatively easy to accommodate, in (21), that the other guests have newspapers.

(21) [Philip goes out for breakfast alone. Nobody is talking to him. But he doesn't care,...]
da er auch eine Zeitung hat
since he also a newspaper has
"since he has a newspaper, too."

Furthermore, Grubic (forthcoming) observes that the materials of the original experiment varied with respect to the discourse-status of the individual whose property needed to be accommodated. In some of the items, that individual was the subject of the first sentence, which had the effect that the story seemed to be told from his or her point of view. They also varied with respect to the role that the accommodated proposition played in the discourse in which they were embedded. In some of them, as in (19), the presupposed meaning contribution of *auch* increases the coherence of the text. Without it, it might be unclear why glasses would make someone more or less likeable to Hannes. The meaning expressed by *auch* provides motivation for this—Hannes and the classmate have something in common. In other items, the presence of *auch* was not crucial for coherence in this way.

We agree with Grubic (forthcoming)'s reasoning that these properties might influence how easy it is for readers/listeners to identify which presupposition the author/speaker intended to convey (this will be discussed in more detail in section 4). We thus also aim to test the following hypothesis:

[6] As no other individual was introduced in the context, it was not possible to formulate the same question in the neutral context. The question in the neutral context instead concerned the asserted rather than the presupposed content, e.g., "Does Isa wear glasses?" in (20b), and is therefore not immediately relevant for the current discussion.

(22) DISCOURSE HYPOTHESIS: Whether a salient individual is sufficient for the felicitous use of *auch* depends on the role of *auch*'s presupposition in discourse.

In other words, we hypothesize that it might be misleading to assess the predictions of the SALIENT INDIVIDUAL and the SALIENT PROPOSITION account based on average judgments across all materials; each account might each make correct predictions for a subpart of the data, depending on the properties of the discourse. Statistically, this would amount to a three-way interaction between TRIGGER TYPE, CONTEXT, and DISCOURSE TYPE. We will describe below in more detail how we controlled for discourse type in our materials.

Finally, great care was taken to ensure that the corresponding sentences without *auch* are entirely felicitous in the neutral context, i.e., that the context and test sentence always form a coherent text, in order to be certain that any infelicity found in these cases is actually due to the additive particle.

3.1.2. Design and Materials

Experiment 1 is based on Grubic (forthcoming)'s design and materials, but with some adjustments. The first difference is that we presented the materials auditorily in order to ensure that the participants correctly identify the associate of the additive particle. Details about the recordings will be provided below.

In our experiment 1, we manipulated the items more systematically with respect to the discourse factors mentioned above by including DISCOURSE TYPE as a between-item factor. The three discourse types are illustrated schematically in (23) and by the examples in (24)–(26). A comparison of the discourse types in (23a) and (23b) shows whether the point of view plays a role, whereas a comparison of the discourse types in (23a) and (23c) shows the influence of coherence. Only the mixed context is shown here, but just like in the written experiment, a positive and a neutral context were constructed as well. Note that the discourse type manipulation does not apply to the neutral context—there, accommodating the presupposition never increased the coherence of the text, and there was always only one individual from whose point of view the story could be construed. In the positive context, on the other hand, the presupposed information is already provided, so the discourse hypothesis does not predict an effect there, either. It specifically predicts an effect on only the mixed context.

(23) a. NON-FACILITATING DISCOURSE:
   (i) the individual relevant for the presupposition does not correspond to the point of view of the text
   (ii) the presupposition is not essential for coherence.
b. FACILITATING PERSPECTIVE:
   (i) the individual relevant for the presupposition corresponds to the point of view of the text
   (ii) the presupposition is not essential for coherence.
c. FACILITATING COHERENCE:
   (i) the individual relevant for the presupposition does not correspond to the point of view of the text
   (ii) the accommodated proposition is essential for coherence

An example item of each type is shown below[7]. Sentence stress is marked by small caps.

(24) *Non-facilitating discourse*:
Mixed context: Paula macht heute einen winterlichen Ausflug mit ihrem Sohn. Sie achtet darauf, auf verschneiten Wegen zu bleiben, …
"Paula is going on a winter excursion with her son today. She is taking care to stay on snow-covered paths, …"
(Paula's point of view)

a. damit sein SCHLITTEN nicht zerkratzt wird.
so.that his sled not scratched is
"so that his sled does not get scratched."

b. weil sie AUCH einen Schlitten hat.
because she also a sled has.
"because she has a sled, too."
(To accommodate: Paula's son has a sled)

(25) *Facilitating perspective:*
Mixed context: Martin will heute etwas Besonderes mit seiner Schwester Lena unternehmen…
"Martin wants to do something special with his sister Lena..."
(Martin's point of view)

a. ...um seinen MASTERABSCHLUSS gebührend zu feiern.
in.order his master's.degree properly to celebrate
"…to celebrate his master's degree."

b. ...um zu feiern, dass sie jetzt AUCH einen Masterabschluss hat.
in.order to celebrate that she now also a master's.degree has
"…to celebrate that she has a master's degree now, too."
(To accommodate: Martin has a Master's degree.)

(26) *Facilitating coherence*:
Mixed context: Alfred mochte seine neue Arbeitskollegin eigentlich ganz gerne als sie bei ihnen angefangen hat. Jetzt sind sie Rivalen, …
"Alfred liked his new colleague when she started. Now they are rivals…"

a. ...weil er ihre AMBITIONEN AUF DEN CHEFSESSEL durchkreuzen will.
because he her ambitions on the executive.chair thwart wants
"…because he wants to thwart her ambitions to become the boss."

[7]The complete list of items for both experiments, as well as our recordings, can be found in the **Supplementary Materials** for this article.

b. ...weil er AUCH Ambitionen auf den
because he also ambitions on the
Chefsessel hat.
executive.chair has
"...because he has ambitions to become the boss, too."
(To accommodate: the colleague is also ambitious — this explains why they are rivals.)

The stimuli were recorded by the first author of the paper with a prosodic realization that unambiguously signaled association of the focus-sensitive *auch* with the subject. This was achieved by assigning sentence stress to *auch* itself. Stressed *auch* associates with a preceding constituent. If several constituents are available, disambiguation is possible by marking one of them as a contrastive topic by a rising accent (Krifka, 1999). In our items, the only available constituent that the particle could associate with is the subject. Association with the direct object would require sentence stress on the object. In the conditions with a possessive pronoun, the object DP (e.g., *ihre Ambitionen auf den Chefsessel* "her ambitions to become the boss" in (26)) was deaccented in the positive context, in which it was explicitly mentioned in the preceding context, and accented in the neutral and mixed contexts, in which it was not prementioned.

3.1.3. Participants and Procedure

Thirty-six native speakers of German took part in the experiment. All of them were students at the University of Potsdam and received credit for participation.

The stimuli were presented using the online questionnaire software SoSci Survey (Leiner, 2018). On the first page, instructions were presented. On each following page of the questionnaire, an audio file was played automatically. At the same time, two questions appeared on the screen. First, the participants were asked to rate the last sentence of the text that they had heard on a scale from 1 (labeled as "very bad, does not make sense in this context") to 5 ("very good, makes sense in the context"). The second task was to answer a question about the content of the text by choosing one of the options "yes," "no," or "I don't know." As in Grubic (forthcoming)'s experiment, this question only targeted the presupposition of the additive particle in the mixed context. The same question was asked in the positive context, where the answer was already provided by the context. In the neutral context, a different question was asked. It was possible to listen to the audio file again if required (the number of replays was not limited). After answering the questions, the participants clicked a button to proceed to the next stimulus. The 18 items were distributed using a Latin Square design and randomized. There were six lists, resulting from crossing the factor TRIGGER TYPE (two levels: possessive vs. additive) with CONTEXT (three levels: mixed vs. neutral vs. positive). DISCOURSE TYPE was included as a between-item factor: six of the items involved a non-facilitating discourse, six involved facilitating perspective, and six involved facilitating coherence. Each participant rated 38 stimuli (two practice trials, 18 critical items and 18 fillers). This way, we obtained 36 data-points for each combination of TRIGGER TYPE/CONTEXT/DISCOURSE TYPE (one from each participant).

3.1.4. Results

The results (collapsing the three levels of the factor DISCOURSE TYPE) are presented in **Figure 3**.

Figures 4, **5** show the results split by DISCOURSE TYPE.

For the statistical analysis, the factor TRIGGER TYPE (additive vs. possessive) was sum-coded. The factor CONTEXT was treatment-coded with the mixed context as the baseline, allowing to compare mixed vs. neutral and mixed vs. positive context. DISCOURSE TYPE was treatment-coded with the non-facilitating discourse as the baseline, allowing to compare non-facilitating vs. facilitating perspective and non-facilitating vs. facilitating coherence.

For distinguishing between the SALIENT PROPOSITION and the SALIENT INDIVIDUAL account with respect to the felicity ratings, it is crucial whether there is an interaction between TRIGGER TYPE and CONTEXT. According to a linear mixed model[8], there was a significant interaction between these factors (mixed vs. neutral context: $t = -3.89, p < 0.001$).

For testing our DISCOURSE HYPOTHESIS, it is crucial to test whether there is an interaction between TRIGGER TYPE, CONTEXT, and DISCOURSE TYPE. No significant three-way interaction was found. The only significant interactions involving DISCOURSE TYPE were a two-way interaction with TRIGGER TYPE (non-facilitating vs. facilitating coherence: $t = -2.56, p = 0.01$) and a two-way interaction with CONTEXT (mixed vs. positive context, non-facilitating discourse vs. facilitating perspective: $t = -2.12, p = 0.03$).

Prior to our analysis of the responses to the question, we collapsed the "no" and "I don't know" answers to a single category (we interpret both answers as indicating non-accommodation of the target proposition). We only included the mixed context in the statistical analysis. In the positive context, the proportion of "yes" answers was at or close to 100%, making a meaningful statistical analysis difficult due to complete separation. In the neutral contexts, a different question was asked (unrelated to accommodation). The remaining factors (TRIGGER TYPE and DISCOURSE TYPE) were included as fixed factors in a logistic regression model. A significant effect of TRIGGER TYPE ($z = 3.09, p = 0.002$) and DISCOURSE TYPE ($z = 2.09, p = 0.04$) was found, but no significant interaction.

3.1.5. Discussion

The direction of the interactions between TRIGGER TYPE and CONTEXT is partially compatible with the predictions of the SALIENT PROPOSITION account: in the mixed context,

[8]The linear models reported in this paper were fit following the recommendations for identifying parsimonious models by Bates et al. (2015a); i.e., we successively reduced the maximal model by removing terms from the random effect structure that showed signs of overfitting until arriving at a model whose principle components all explain non-zero variance and which provides a better fit than the minimal model (including only random intercepts). We used the R packages lme4, lmerTest, and RePsychLing (Baayen et al., 2015; Bates et al., 2015b; R Core Team, 2016; Kuznetsova et al., 2017). The full model specifications and results are provided as **Supplementary Materials**. Here, we only report the results that are directly relevant for the tested hypotheses.

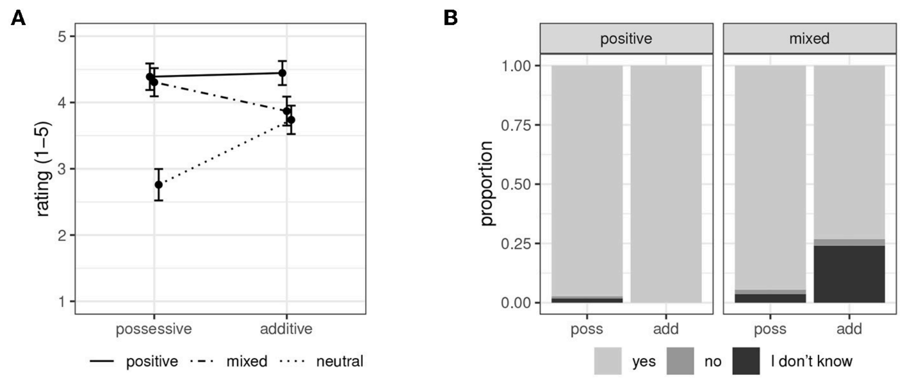


FIGURE 3 | Results of experiment 1: felicity ratings with 95% confidence intervals **(A)** and answer proportions **(B)**.

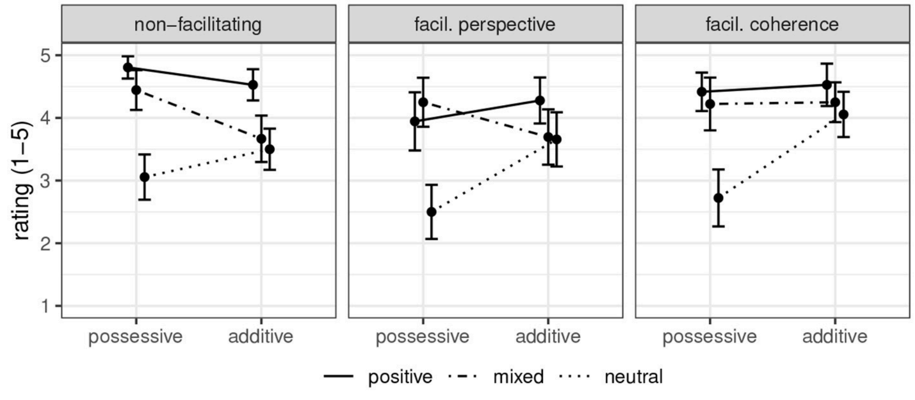


FIGURE 4 | Felicity rating results of experiment 1 with 95% confidence intervals, split by discourse type.

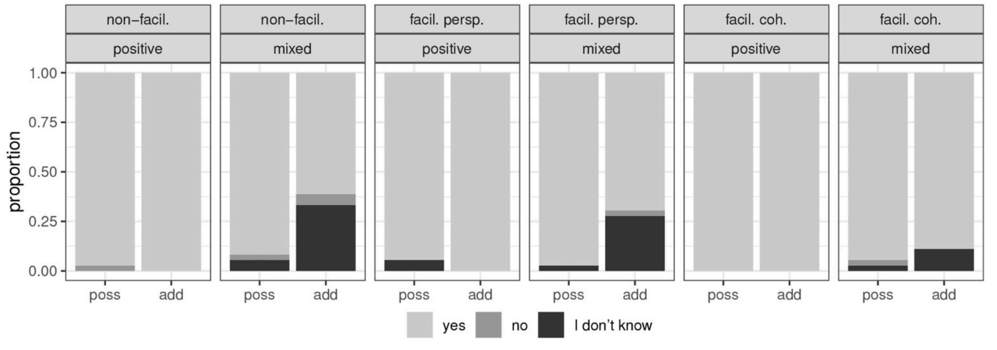


FIGURE 5 | Question proportion results of experiment 1, split by discourse type.

additives are less acceptable than possessives (more precisely, the felicity difference between the mixed and the neutral context is smaller for additives than for possessives). However, as in Grubic's (forthcoming) original study, the interaction is driven by the high acceptability of additives in the neutral context rather than by low acceptability in the mixed context.

The results also show, and this is not predicted by the SALIENT PROPOSITION account, that DISCOURSE TYPE significantly affects the felicity of additives. The conditions including an additive are generally more acceptable (in comparison to possessives) in the facilitating coherence discourse type than in the non-facilitating discourse type, even though the proposition to be accommodated is still not entailed by the context, but instead merely becomes easier to identify because it enhances the coherence between the context and the test sentence. Note however, that although DISCOURSE TYPE should only have an effect in mixed contexts, the judgments of additives in the neutral context were also enhanced. This is unexpected because the intended facilitating perspective/coherence was not present in the neutral context. Thus, our DISCOURSE HYPOTHESIS (predicting more specific effects for the different context rather than an overall facilitation) was not confirmed.

Since DISCOURSE TYPE was a between-items factor in our design, it is possible that another aspect in which the items differed lead to the differences in the neutral context; we will address this possibility in the second experiment.

Our conjecture that the high ratings in the neutral context might arise because participants assign a different implicit prosody than intended was not confirmed: change of modality from written to auditory stimuli did not change this. However, it is conceivable that participants did not pay attention to the prosody in the items. In order to be certain that the associate of the additive particle was identified correctly by the participants, we took further measures in Experiment 2 (reported in the next section) to rule out any misinterpretation of the test sentences.

A second result not predicted by the SALIENT PROPOSITION account is that, like in the experiment reported by Grubic (forthcoming), the proportion of "yes" answers to the content question was higher for possessives than for additives, but for both trigger types, the "no"/"I don't know" answers are clearly in the minority. This suggests that most participants accommodated the target proposition in the mixed context. Numerically, there is a trend toward more accommodation in the facilitating coherence discourse type.

Based on Experiment 1, we tentatively conclude that the SALIENT PROPOSITION account, predicting that additives are only licit when the presupposition is entailed by the context, is too strong—at least in facilitating discourse types, the additive particle is acceptable to a similar degree in the mixed context as in the positive context. The SALIENT INDIVIDUAL account, predicting that additives should be felicitous whenever there is another salient individual, is too weak—in the absence of facilitating discourse factors, the acceptability of the additive particle is similarly low in the mixed context as in the neutral context. However, conclusive assessment of the accounts is impeded by the unpredicted behavior of additives in the neutral context.

3.2. Experiment 2

3.2.1. Motivation

In Experiment 1 discussed above, the result that additive particles were judged to be felicitous in a context that does not support their presupposition, which was already found in the experiment reported in Grubic (forthcoming), was corroborated. This is unpredicted considering the previous literature, which states that the presupposition of additive particles is hard to accommodate.

Assuming that the presupposition of additive particles is not in fact freely accommodatable, there are two possible explanations for this result. First, even though in the auditory stimuli main stress was placed on the additive particle, indicating association with the subject, the participants might not have paid ample attention to the intonation. Since the word order also allows for association of *auch* with the VP, this would lead to a presupposition that the subject of the sentence has some other property, which is satisfied in most if not all or our neutral contexts, and would explain the high felicity judgments. For example, in (27), the context entails that Jakob has several properties (being a student, talking to Klara), so the presupposition that he has a further property apart from having a sister is satisfied in this context.

(27) Klara unterhält sich mit ihrem Kommilitonen Jakob darüber, wie es ist, auf dem Land aufzuwachsen. Es stellt sich heraus, …
"Klara is talking to her fellow student Jakob about growing up on the countryside. It turns out … "
dass Jakob auch eine Schwester hat.
that Jakob also a sister has
"that Jakob has a sister, too."

A second possible explanation was that the possessum, e.g., a sister in (27), was usually something which generally a lot of people have (e.g., a sister, a car, a TV, i.a.). It is a central claim of the literature on the anaphoricity of *too* that world knowledge indicating that many people have the property in question does not license the use of the additive particle. For example, in Kripke's example (28), the knowledge that a lot of people are having dinner in New York on any given night does not, according to Kripke, make *too* felicitous (Kripke, 2009, p. 373). Nevertheless, we wanted to follow up on the hypothesis that the participants accommodate something of this sort, e.g., in example (27) that many other people have sisters.

(28) Sam is having dinner in New York tonight, too
(To accommodate: Somebody else is having dinner in New York tonight)

Since this factor was not controlled for systematically in Experiment 1, it might also have lead to differences between the items, and, as a consequence, confounded the between-items factor DISCOURSE TYPE.

3.2.2. Design and Materials

The design and materials for Experiment 2 were based on Experiment 1, but the following three changes were made: first, all direct objects referring to something that people commonly have were replaced by more specific expressions, referring to something that most people do not have, but is nevertheless relatively uncontroversial[9]. For example, an object like *sister* was replaced by the more specific *twin sister*, as illustrated in (29).

[9]That this item is something uncontroversial is important in order to enable accommodation of the presupposition of possessives in mixed contexts in our study (i.e., those that do not entail the possession relation). For example, the presupposition that the speaker owns a cat can easily be accommodated in (ia), but the corresponding presupposition that she owns a gorilla in (ib) is not easy to accommodate, because it is not common to own a gorilla (Krifka, 2008, p. 246).

(i) a. I had to bring **my cat** to the vet because it was sick.
b. I had to bring **my gorilla** to the vet because it was sick.

(29) *Adjusted item from Experiment 2:* Klara unterhält sich mit ihrem Kommilitonen Jakob darüber, wie es ist, auf dem Land aufzuwachsen. Es stellt sich heraus, …
"Klara is talking to her fellow student Jakob about growing up on the countryside. It turns out … "

a. dass Jakob AUCH eine Zwillingsschwester hat.
that Jakob also a twin=sister has
"that Jakob has a twin sister, too."

b. dass Jakob IHRE ZWILLINGSSCHWESTER kennt.
that Jakob her twin=sister knows
"that Jakob knows her twin sister."

The second change was the addition of a third type of test sentence. In order to exclude the possibility that participants interpreted the additive particle as associating with something other than the subject (in spite of the prosodic cues), we tested a further type of structure in which the interpretation is made unambiguous by both prosody and syntax. In a sentence like (30), the additive particle can only associate with the subject it precedes.

(29) c. dass auch JAKOB eine Zwillingsschwester hat.
that also Jakob a twin=sister has
"that Jakob too has a twin sister."

The third change was made to the follow-up questions in the neutral context. In Experiment 1, these follow-up questions were about the assertion of the test sentence (e.g., "Does Jakob have a sister?") in the neutral context and thus did not help decide whether participants accommodate the correct additive presupposition. The questions for the neutral context were thus changed in Experiment 2, so that they target the presupposition instead (e.g., "Does somebody else have a twin sister?").

3.2.3. Participants and Procedure

Fifty-four native speakers of German took part in the experiment (resulting in the same number of data points per condition as for Experiment 1). Participants were recruited via Prolific. The stimuli were presented using the online questionnaire software L-Rex (Starschenko, 2018). The procedure was the same as described for experiment 1. Again, each participant rated 38 stimuli (two practice trials, 18 critical items and 18 fillers).

3.2.4. Results

The results are presented in **Figures 6–8**. **Figure 6** (again collapsing the three levels of DISCOURSE TYPE) presents the mean felicity ratings and the answer proportions.

The results for the three different discourse types are presented in **Figures 7**, **8**.

For the statistical analysis, the factors CONTEXT and DISCOURSE TYPE were again treatment-coded, as described for Experiment 1. The factor TRIGGER TYPE was Helmert-coded to allow comparison between additive (preposed/postposed) vs. possessive, and additive preposed vs. additive postposed.

With respect to the felicity ratings, according to a linear mixed model the factor TRIGGER TYPE (additive vs. possessive) interacted significantly with CONTEXT (mixed vs. positive context: $t = -3.82$, $p < 0.001$, mixed vs. neutral context: $t = -4.81, p < 0.001$), as in Experiment 1.

However, in contrast to the first experiment, this two-way interaction was qualified by higher-order interactions with DISCOURSE TYPE. The DISCOURSE HYPOTHESIS states that the strength of the TRIGGER TYPE (additive vs. possessive) × CONTEXT interaction differs depending on DISCOURSE TYPE; but it does not specify which levels of CONTEXT and DISCOURSE TYPE should be affected (because it was not possible to derive more specific hypotheses from the previously available data). There are therefore four possible interaction terms that would lend support to the discourse hypothesis if they were significant. We correct for the higher probability of Type I error in multiple testing using the Holm-Bonferroni method. The interaction with the lowest unadjusted p-value (additive vs. possessive; mixed vs. positive context; non-facilitating vs. facilitating coherence: $t = 2.59, p = 0.0097$) was significant at the adjusted α level of 0.0125; the others were not[10].

Only one marginally significant difference was found between pre-posed and post-posed *auch* in form of an interaction with discourse type (non-facilitating vs. facilitating perspective: $t = -1.89, p = 0.059$).

Again, the "no" and "I don't know" responses were collapsed for the statistical analysis, and only the mixed context was included. TRIGGER TYPE (additive vs. possessive) interacted significantly with DISCOURSE TYPE (non-facilitating vs. facilitating coherence: $t = -2.65$, $p = 0.008$). A marginally significant overall difference between pre-posed and post-posed *auch* was found ($t = 1.89, 0.059$).

3.2.5. Discussion

One of the motivations for this experiment was to test whether participants misinterpreted the additive test sentences in Experiment 1 by choosing the wrong associate for postposed *auch*.

The felicity rating results indicate that there was little or no such misinterpretation: only marginally significant differences between postposed *auch* and preposed *auch* (which unambiguously associates with the subject in our test sentences) were found, and the crucial observations with respect to the felicity ratings—that the mixed context is closer to the neutral context for additives, and closer to the positive context for possessives—hold for both versions of *auch*. Nevertheless, the trend toward less "yes" and more "I don't know" answers for postposed *auch* than for preposed *auch*, especially in the non-facilitating discourse type, could be tentatively interpreted as an indication that participants felt more certain about the accommodated presuppostion when the associate for *auch* was marked syntactically, not only prosodically, and that it would therefore be advisable to use preposed *auch* in future research.

Generally, the results of Experiment 1 were replicated. First, an interaction between TRIGGER TYPE and CONTEXT was found, as predicted by the SALIENT PROPOSITION account (the mixed

[10]Second-lowest unadjusted p-value: additive vs. possessive; mixed vs. neutral context; non-facilitating vs. facilitating perspective: $t = 2.04, p = 0.042$, n.s. at the adjusted α level.

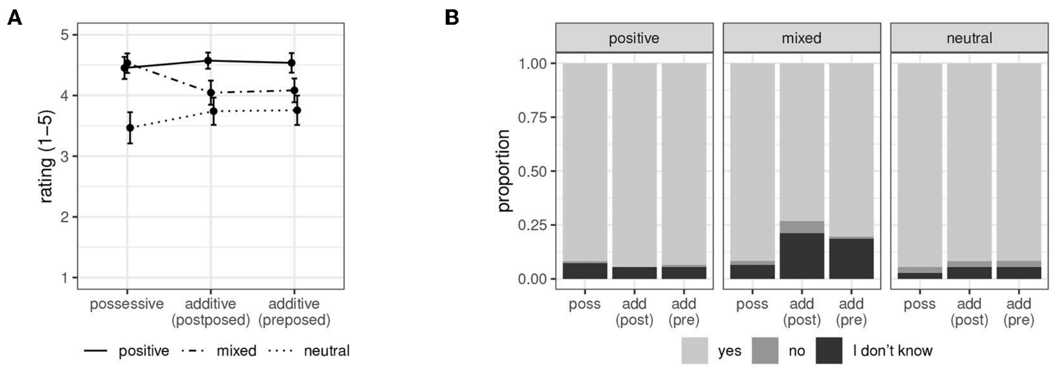


FIGURE 6 | Results of Experiment 2: felicity ratings with 95% confidence intervals **(A)** and answer proportions **(B)**.

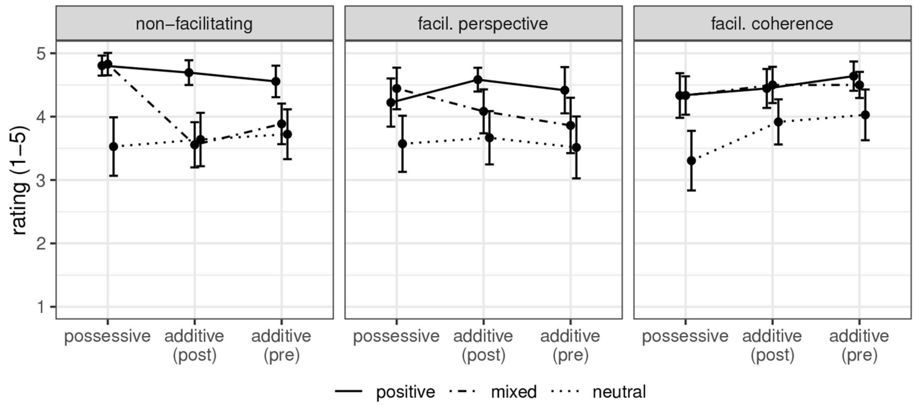


FIGURE 7 | Felicity rating results of Experiment 2 with 95% confidence intervals, split by discourse type.

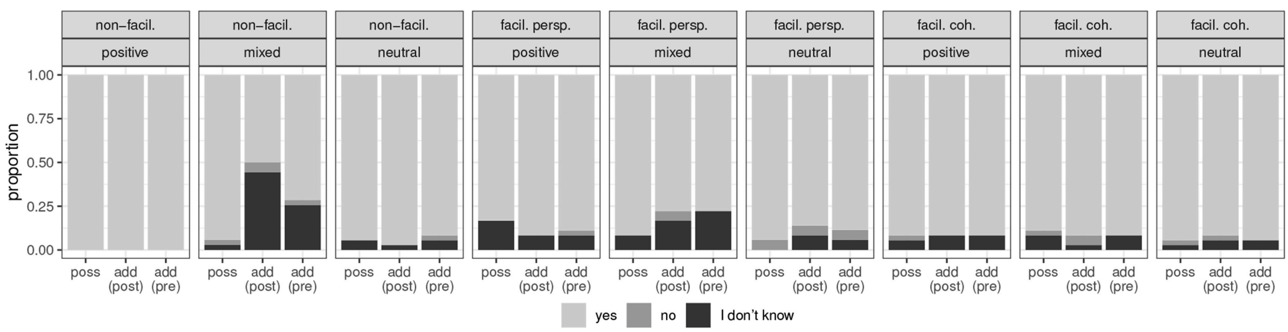


FIGURE 8 | Question proportion results of Experiment 2, split by discourse type.

context is closer to the neutral one for additives than for possessives, and closer to the positive context for possessives than for additives), but not the SALIENT INDIVIDUAL account.

But again, as not predicted by the SALIENT PROPOSITION account, DISCOURSE TYPE was found to play a role for the felicity judgments. Like in experiment 1, the facilitating coherence discourse type again raised the ratings of the items with additives. But in contrast to Experiment 1, this did not affect the mixed and neutral context in the same way: it specifically raised the felicity of additives in the mixed context (as evidenced by the three-way interaction). While in the non-facilitating discourse type, the felicity ratings of the additive particles in the mixed context are very close to the negative baseline, they are very close to the positive baseline in the facilitating coherence discourse type. The fact that discourse type had a more specific effect in Experiment 2 rather than generally raising the acceptability (even in the neutral context) could be due to our effort to make the items more uniform with respect to how common the relevant objects were; this might have eliminated a confound and therefore reduced some of the random noise between items (and thus, crucially, between the discourse types).

When we compare the results to the predictions that were presented in **Figure 1**, it is particularly striking that the pattern in the non-facilitating discourse type corresponds almost exactly to the predictions of the SALIENT PROPOSITION account, whereas the pattern in the facilitating coherence discourse type is similar to the predictions of the SALIENT INDIVIDUAL account[11] (except for the fact that the ratings in the neutral context are still somewhat higher for additives than for possessives). Thus, the tension between the two accounts might be resolved by taking into account discourse properties which can facilitate the identification of the proposition that needs to be accommodated.

In sum, these findings further support the view that a salient individual is not sufficient to make the use of the additive particle felicitous. A salient proposition, on the other hand, is not a necessary condition for felicity, at least in the sense that it does not necessarily need to be present in the immediately preceding context. Accommodation is possible and facilitated by discourse factors. In the facilitating coherence discourse type, an additive can be as felicitous in the mixed context as in the positive context.

Although controlling the commonness/frequency of the possessum more systematically in Experiment 2 indeed helped to reduce some of the noise in the data and to make possessives and additives more comparable in the neutral context, our hypothesis that the surprisingly high felicity ratings for additive particles in this context might be due to this factor was not fully confirmed. What we observe is an increase in the felicity of the possessives, not a decrease in the felicity of the additives in comparison to Experiment 1. Inspection of the fillers suggests that this is not due to a scale bias, i.e., generally lower ratings by the participants of Experiment 2: on average, the fillers were rated a little bit higher than in Experiment 1 (3.86 vs. 3.66), but not to an extent that could fully explain the difference in the felicity of the possessives in the neutral context between the experiments. It is an open question whether this was caused by the choice of more specific objects, and if so, why it affected the felicity of the possessives in this way. We do however have some hypotheses as to why, in our experiments, additive particles received substantially higher felicity ratings in neutral contexts than in Tiemann (2014)'s experiment (e.g., (30)) and than our own filler item (9) (repeated here as (31)).

[11]This is supported by a *post-hoc* analysis of these two subsets of the data: in the non-facilitating discourse type, there is a significant interaction between TRIGGER TYPE (additive vs. possessive) and CONTEXT both for mixed vs. positive ($t = -4.88, p < 0.001$) and mixed vs. neutral ($t = -4.22, p < 0.001$) context, in line with the predictions of the SALIENT PROPOSITION account. In the facilitating coherence discourse type, there is no significant difference between the positive and the mixed context (neither as a main effect nor in interaction with TRIGGER TYPE. For the neutral context, as predicted by the SALIENT INDIVIDUAL account, a significant main effect was found in comparison to the mixed context ($t = -2.88, p = 0.03$), but also a significant interaction with TRIGGER TYPE ($t = 2.00, p = 0.048$), due to the still higher than expected values of the additives in the neutral context.

(30) Nobody is eating soup with Tina today.
Sie hofft, dass auch Susanne eine Suppe mit ihr kocht
She hopes that too Susanne a soup with her cooks
und kauft dafür Zutaten
and buys for.it ingredients
"She hopes that Susanne will cook a soup with her, too, and buys ingredients for it."

(31) Mia is eating a salad on the bus. A stranger sits down next to her and says:
Unser Busfahrer isst auch ein Brötchen.
our bus.driver eats also a bun
"Our bus driver is eating a bun, too"

We believe that part of the reason is, first, that we took great care to ensure that the discourse is entirely natural and coherent without the particle, so that any infelicity stems from additive particle itself. Second, our neutral contexts did not contain any propositions that are similar to a potential antecedent. We hypothesize that (31) might be particularly infelicitous because the participants try to accommodate *Mia is eating a salad* as the required antecedent for *our busdriver is eating a bun*, and fail.

As a final methodological remark about our experiments, we would like to point out that our data shows a correlation between whether the answer to the content question indicated accommodation and the felicity rating[12]. This provides support for the view that felicity judgments can be used to investigate presupposition accommodation: when a presupposition cannot be accommodated, the test sentence receives a lower felicity rating.

4. CONCLUSION

4.1. Summary

This paper reports the results of two experiments aiming to test the salience requirements of the German additive particle *auch*. We tested the predictions made by the two main accounts on the anaphoricity of additive particles found in the literature: the SALIENT INDIVIDUAL account, which predicts that *auch* just requires a focus alternative to the focused constituent to be salient, and the SALIENT PROPOSITION account, which predicts that the context needs to entail a focus alternative to the entire proposition containing *auch*. Sentences with *auch* were thus tested in three different contexts: ones which entailed the proposition (positive context), ones which merely made an individual salient (mixed context), and ones which neither made a proposition nor an individual salient (neutral context).

Our results suggest that the predictions of the SALIENT INDIVIDUAL account (e.g., Heim, 1992; Geurts and van der Sandt, 2004) are too weak: when controlling for independent accommodation-enhancing factors like perspective and coherence, the existence of a salient individual in the context

[12]For additives in the mixed context, higher ratings were found in trials in which the question was answered by "yes" than in trials in which the question was answered by "no" or "I don't know" in both experiments (experiment1: 4.09 vs. 3.28; significant according to a *t*-test: $t = 3.08, p = 0.004$; experiment 2: 4.28 vs. 3.34; significant according to a *t*-test: $t = 5.93, p < 0.001$).

was not enough to make *auch* felicitous. However, our results also suggest that a strict version of the SALIENT PROPOSITION account is too strong: when the context does not entail the proposition, but enhances the probability that a certain prementioned individual is in fact the one relevant for the presupposition, *auch* becomes more felicitous. Two such accommodation-enhancing discourse types were tested: (i) when the preceding discourse can be seen to be from the point of view of the individual that the presupposition is about (perspective), and (ii) when the presupposition makes the text more coherent (coherence). Our experiments provide evidence for a significant effect of coherence (found for the judgments in Experiments 1 and 2, and for the proportion of answers indicating successful accommodation in Experiment 2). In the facilitating coherence discourse type, *auch* was judged to be as felicitous in the mixed context as in the positive context, and the proposition was indicated to be true to almost the same extent in the two types of contexts. These results thus clearly provide evidence against a SALIENT PROPOSITION account which requires that the proposition be entailed by the immediate linguistic or non-linguistic context (e.g., Ruys's 2015 Givenness account). Concerning other variants of the SALIENT PROPOSITION account, it depends on how exactly they define salience. As indicated by previous experimental findings (see section 2.1), hearers use previous linguistic material in order to identify the correct propositional focus alternative relevant for the presupposition. In the facilitating coherence cases, the relevant propositional focus alternative can be identified not because it is prementioned, but because that particular information makes the text coherent. For example, in (25), repeated here, the presupposed information ("Alfred's colleague has ambitions to become the boss") makes the causal relationship between the main clause ("Alfred and his colleague are rivals") and the embedded clause ("because Alfred has ambitions to become the boss") more coherent, based on a previous assumption that in order to be rivals, two people have to have a common ambition[13].

(26) FACILITATING COHERENCE:
Alfred mochte seine neue Arbeitskollegin eigentlich ganz gerne als sie bei ihnen angefangen hat. Jetzt sind sie Rivalen, …
"Alfred liked his new colleague when she started. Now they are rivals…"
...weil er AUCH Ambitionen auf den Chefsessel hat.
because he also ambitions on the executive.chair has
"…because he has ambitions to become the boss, too."
(To accommodate: the colleague is also ambitious — this explains why they are rivals.)

In order to account for these data, the relevant notion of salience would thus have to be one that also applies to inferences arising from the utterance itself, together with its context.

4.2. Accommodatability: The Bigger Picture

In this section, we briefly discuss the question whether our results have a bearing on the question what differentiates easily accommodatable from less easily accommodatable presuppositions. In the previous literature, three main reasons are proposed why certain triggers have presuppositions that are hard to accommodate: First, it has been proposed that these triggers presuppose that something is "in the discourse record" (see e.g., the proposal in Beaver and Zeevat, 2007, for *too*), i.e., immediately prementioned or mutually attended to and thus, according to Beaver and Zeevat, not accommodatable, see (32).

(32) THE DISCOURSE RECORD PRINCIPLE: Presuppositions about what is in the discourse record may not be accommodated

Second, some authors have suggested that presuppositions which are not crucially important for the truth-conditional meaning of the sentence can be ignored, see (33) (Tiemann 2014; Tiemann et al. 2015, cf. also Zeevat, 2002 and Jäger and Blutner, 2003's "Do Not Accommodate," as discussed in Beaver and Zeevat, 2007)[14]. For additive particles the truth-conditional meaning of the sentence is simply the meaning of the sentence without the particle, the additive presupposition is thus not needed for the truth conditions and can be ignored according to (33).

(33) MINIMIZE ACCOMMODATION: Do not accommodate a presupposition unless missing accommodation will lead to uninterpretability of the assertion!

Third, as noted above, van der Sandt (1992) and Geurts and van der Sandt (2004) propose that presupposition triggers that are not sufficiently semantically rich cannot be accommodated because the addressee cannot identify the presupposition to be accommodated. Beaver and Zeevat (2007) formulate this as in (34).

(34) THE INSUFFICIENT CONTENT PRINCIPLE: Accommodation is only possible when the presupposition is descriptively rich. If a low content presupposition cannot be resolved, infelicity results.

Even though, on the surface, additive particles give rise to a presupposition that appears to be rich in content, Geurts and van der Sandt (2004)—as noted above—assume that they involve two presuppositions, of which one is similar to a pronoun and thus descriptively poor and not accommodatable.

[13] Note that non-truthconditional meaning is usually not seen to be part of this causal relationship (Beaver and Clark, 2008, p.217), e.g., in (i), there is no inference that Karl asks the waiter for a sharp knife (partly) because Anna has a steak.

(i) Karl is in a restaurant with Anna. Anna has ordered a steak. When their food is brought to them, Karl asks the waiter for a sharp knife, because
er auch ein Steak hat.
he also a steak has
"he has a steak, too."

[14] For experimental results suggesting such differences see Cummins et al. (2012), Domaneschi et al. (2014), and Amaral and Cummins (2015).

Our results do not support any hypothesis which strictly differentiates between two categories of triggers, on the basis of some part of their presupposition which is (un)accommodatable (e.g., the DISCOURSE RECORD PRINCIPLE and a strict understanding of the INSUFFICIENT CONTENT PRINCIPLE). In addition, we do not believe that MINIMIZE ACCOMMODATION properly explains the difference found here. It is true that in the facilitating coherence case, the presupposition is—in a sense—more relevant for the truth conditions because it is needed to establish coherence. However, MINIMIZE ACCOMMODATION does not account for the surprising level of acceptability that we find with *auch* sentences in general, nor does it, in our view, give credit to the effort that addressees make to identify a proposition which they can accommodate as the presupposition of *auch*. We suggest, instead, that our results support an account of *identifiability* (a weak version of the INSUFFICIENT CONTENT PRINCIPLE), which might be formulated as in (35)[15].

(35) IDENTIFIABILITY: Presuppositions can only be accommodated when the addressee can identify what exactly to accommodate.

In the case of *auch*, the individual relevant for the presupposition has to be identified, while the remainder of the presupposition can be deduced from the utterance containing the additive particle. Our results showed that while previous mention of an individual was not sufficient in order to allow for accommodation, a further indication that this individual is in fact the required focus alternative facilitated the accommodation of the additive presupposition. Further evidence comes from examples where the alternative set only contains two relevant alternatives, e.g., (36)[16]. There, no special context is needed in order to license the use of *too*, since the identity of the other alternative, being the only other focus alternative in the alternative set, is already clear.

(36) Bereavement benefits must support unmarried couples too.
(To accommodate: bereavement benefits support married couples.)

Thus, while we hypothesize that coherence can generally be used to help identify the information to be accommodated (not only in the case of additive particles), there are other mechanisms, such as the role of a restricted alternative set as in (36), which are limited to focus-sensitive particles[17].

4.3. Outlook: *auch* in Neutral Contexts

One result of our experiments was that, just like in Grubic's (forthcoming) experiment, additive particles were not as bad as predicted in neutral contexts. We showed that this was not due to a misunderstanding on the part of the experiment participants: they were able to correctly identify the associate of *auch*.

In the experiments reported here, we wanted to exclude any factors which might enhance the acceptability of additive particles in the neutral context, since we were mainly interested in factors enhancing their acceptability in the mixed context. For this reason we excluded, for the first experiment, any items in which we believed that the restrictor of a quantifier in the context might provide a kind of implicit domain for the additive particle (e.g., "nobody [of the other guests] is talking to him, but he doesn't mind because he has a newspaper, too"). We also excluded any items that are frequent possessions, in order to prevent potential accommodation of a very weak presupposition (e.g., "people in general have TVs").

It however remains an interesting question what is responsible for the relatively high acceptability of additives in neutral contexts. Two anonymous reviewers pointed out two hypotheses worth testing. The first is whether participants accommodate that the speaker (or addressee) is the individual relevant for the presupposition (as e.g., in "Are you presenting at SALT, too?"). This might e.g., play a role in (37), where we readily assume that the person trying to reach Jan is the speaker.

(37) *Neutral context:* Jan ist telefonisch nicht mehr zu erreichen. Es stellt sich heraus, "Jan is not reachable via phone anymore. It turns out"
dass er auch eine neue Nummer hat.
that he also a new number has
"that he has a new number, too."

The second hypothesis is that implicit domains may play a role in more cases than we identified. This very likely plays a role for the acceptability of (38). Here, since *the best group* is mentioned, it is easy to accommodate that all individuals in this group have a good style of drawing[18].

(38) *Neutral context:* Gordon geht zum Einstufungstest für einen Kunstkurs. Er wird in die beste Gruppe eingeteilt, "Gordon participates in an entry-level test for an art course. He gets assigned to the best group"
weil er auch einen guten Zeichenstil hat.
because he also a good drawing-style has
"because he also has a good style of drawing."

[15]Bacovcin et al. (2018) offer a proposal based on Sudo (2012) and Klinedinst (2016) which, in a sense, also suggests that identifiability is a key factor: some triggers also entail their presupposition, which is then easier to accommodate than non-entailed presuppositions, since addressees can rely on this entailed meaning in order to identify the presupposition—we thank an anonymous reviewer for pointing out this reference to us. Zehr and Schwarz (2016) show that the additive presupposition is not entailed.

[16]An anonymous reviewer pointed out that example (36) is most natural with a focus accent on "un," which seems to corroborate our claim that there are only two considered focus alternatives here.

[17]A reviewer asks whether we assume that coherence generally influences inferences made by the hearer. It might well be that implicatures might be strengthened if they help make the utterance that gives rise to them coherent, but this is only speculation on our part, and needs to be shown in further research.

[18]While this is a very clear example, the reviewer suggests that something like this may happen even if just a place where other people may be (an office, a restaurant, i.a.) is mentioned, which is quite frequently—but not always—the case in our neutral context examples.

We agree with these intuitions and believe that they should be tested in further research. Since the important factor for the accommodatability of *auch*'s presupposition is its identifiability, there are presumably many factors at play which help identify a presupposition, and thus enhance its accommodatability.

ETHICS STATEMENT

Since our studies were online linguistic questionnaires involving a healthy adult population, no approval by an ethics committee was required according to institution's guidelines and national regulations. Participants' consent was obtained by virtue of their voluntary participation in our study and their data were fully anonymized.

AUTHOR CONTRIBUTIONS

MG designed and ran the initial experiment that the current experiments are based on, and wrote the majority of the introduction, background and conclusion. MG and MW designed and ran the two experiments together. MW performed the statistical analysis, and wrote the majority of the section on the experiments.

FUNDING

This publication was funded by the Open Access Publication Fund of the University of Potsdam. The Publication Fund is supported by Deutsche Forschungsgemeinschaft (German Research Foundation) within the framework of the Open Access Publishing program. We would like to thank the DFG and the University of Potsdam for funding the publication of this paper.

ACKNOWLEDGMENTS

The authors would like to thank the editors, Sophie Repp and Katharina Spalek, for the possibility to contribute to this research topic, and the editors and two reviewers for their helpful comments. All remaining errors are our own. We would like to thank the DFG and the University of Potsdam for this support. We additionally thank our colleagues at the University of Potsdam, in particular Malte Zimmermann and Mareike Philipp, for valuable comments on our first experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2019.00015/full#supplementary-material

REFERENCES

Amaral, P., and Cummins, C. (2015). "A cross-linguistic study on information backgrounding and presupposition projection," in *Experimental Perspectives on Presuppositions*, ed F. Schwarz (Cham; Heidelberg; New York, NY; Dordrecht; London: Springer), 157–172.

Asher, N., and Lascarides, A. (1998). The semantics and pragmatics of presupposition. *J. Semant.* 15, 239–300.

Baayen, H., Bates, D., Kliegl, R., and Vasishth, S. (2015). *RePsychLing: Data Sets From Psychology and Linguistics Experiments*. R package version 0.0.4.

Bacovcin, H. A., Zehr, J., and Schwarz, F. (2018). To accommodate or to ignore?: the presuppositions of again and continue across contexts. *Glossa J. Gen. Linguist.* 3:16. doi: 10.5334/gjgl.402

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv:1506.04967 [stat.ME]*.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67. doi: 10.18637/jss.v067.i01

Beaver, D., and Clark, B. (2008). *Sense and Sensitivity: How Focus Determines Meaning*. Chichester: Wiley-Blackwell.

Beaver, D., and Zeevat, H. (2007). "Accommodation," in *The Oxford Handbook of Linguistic Interfaces*, eds G. Ramchand and C. Reiss (Oxford: Oxford University Press), 503–538.

Büring, D. (2004). Focus suppositions. *Theor. Linguist.* 30, 65–76. doi: 10.1515/thli.2004.007

Chemla, E., and Schlenker, P. (2012). Incremental vs. symmetric accounts of presupposition projection: an experimental approach. *Nat. Lang. Semantics* 20, 177–226. doi: 10.1007/s11050-012-9080-7

Corblin, F. (1991). *Presupposition and Discourse Context.* Paris: CNRS.

Cummins, C., Amaral, P., and Katsos, N. (2012). Experimental investigations of the typology of presupposition triggers. *Hum. Mente* 23, 1–15.

Domaneschi, F., Carrea, E., Penco, C., and Greco, A. (2014). The cognitive load of presupposition triggers: mandatory and optional repairs in presupposition failure. *Lang. Cogn. Neurosci.* 29, 136–146. doi: 10.1080/01690965.2013.830185

Geurts, B., and van der Sandt, R. (2004). Interpreting focus. *Theor. Linguist.* 30, 1–44. doi: 10.1515/thli.2004.005

Gotzner, N., and Spalek, K. (2014). "Exhaustive inferences and additive presuppositions: the interplay of focus operators and contrastive intonation," in *Proceedings of the ESSLLI Workshop on Formal and Experimental Pragmatics*, eds J. Degen, M. Franke, and N. Goodman (Tübingen), 7–13.

Grubic, M. (forthcoming). "Additives and accommodation," *Secondary Content: The Linguistics of Side Issues*, eds D. Gutzmann and K. Turgay (Leiden: Brill), 169–200.

Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *J. Semant.* 9, 183–221.

Jäger, G. (2004). Alternatives or presuppositions? A comparison of the background presupposition rule with alternative semantics. *Theor. Linguist.* 30, 111–122.

Jäger, G., and Blutner, R. (2003). "Competition and interpretation: the German adverb wieder ("again")," in *Modifying Adjuncts*, eds E. Lang, C. Maienborn, and C. Fabricius-Hansen (Berlin: Mouton de Gruyter), 393–416.

Kim, C. S. (2012). *Generating Alternatives: Interpreting Focus in Discourse*. Ph.D. thesis, University of Rochester.

Kim, C. S. (2015). "Presupposition satisfaction, locality and discourse constituency," in *Experimental Perspectives on Presuppositions*, ed F. Schwarz (Cham; Heidelberg; New York, NY; Dordrecht; London: Springer), 109–134.

Klinedinst, N. (2016). "Two types of semantic presuppositions," in *Pragmemes and Theories of Language Use*, eds K. Allan, A. Capone, and I. Kecskes (Cham: Springer International Publishing), 601–624.

Krifka, M. (1999). "Additive particles under stress," in *Proceedings of SALT 8* (Ithaca, NY: Cornell; CLC Publications), 111–128.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguist. Hungar.* 55, 243–276. doi: 10.1556/ALing.55.2008.3-4.2

Kripke, S. A. (2009). Presupposition and anaphora: remarks on the formulation of the projection problem. *Linguist. Inq.* 40, 367–386. doi: 10.1093/acprof:oso/9780199730155.001.0001

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. *J. Stat. Softw.* 82. doi: 10.18637/jss.v082.i13

Leiner, D. J. (2018). *SoSci Survey (Version 2.6.00-i) [Computer Software]*. Available online at: https://www.soscisurvey.de (accessed April 9, 2019).

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Available online at: https://www.R-project.org

Roberts, C. (2010). *Retrievability and Incomplete Descriptions.* Columbus, OH: The Ohio State University.

Romoli, J., Khan, M., Sudo, Y., and Snedeker, J. (2015). "Resolving temporary referential ambiguity using presupposed content," in *Experimental Perspectives on Presuppositions*, ed F. Schwarz (Cham; Heidelberg; New York, NY; Dordrecht; London: Springer), 67–87.

Rooth, M. (1985). *Association With Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semant.* 1, 75–116.

Rooth, M. (1996). "Focus," in *The Handbook of Contemporary Semantic Theory*, ed S. Lappin (London: Basil Blackwell), 271–297.

Ruys, E. G. (2015). On the anaphoricity of too. *Linguist. Inq.* 46, 343–361. doi: 10.1162/LING_a_00184

Schwarz, F. (2007). Processing presupposed content. *J. Semantics* 24, 373–416. doi: 10.1093/jos/ffm011

Schwarz, F. (2015). "Presuppositions vs. asserted content in online processing," in *Experimental Perspectives on Presuppositions*, ed F. Schwarz (Cham; Heidelberg; New York, NY; Dordrecht; London: Springer), 89–108.

Schwarz, F. (2016). Experimental work in presupposition and presupposition projection. *Annu. Rev. Linguist.* 2, 273–292. doi: 10.1146/annurev-linguistics-011415-040809

Schwarzschild, R. (1999). GIVENness, avoidF and other constraints on the placement of accent. *Nat. Lang. Semant.* 7, 141–177.

Singh, R., Fedorenko, E., Mahowald, K., and Gibson, E. (2016). Accommodating presuppositions is inappropriate in implausible contexts. *Cogn. Sci.* 40, 607–634. doi: 10.1111/cogs.12260

Starschenko, A. (2018). *L-Rex [Computer Software].* Available online at: https://github.com/2e2a/l-rex/ (accessed April 9, 2019).

Sudo, Y. (2012). *On the Semantics of Phi Features on Pronouns*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Tiemann, S. (2014). *The Processing of Wieder ('Again') and Other Presupposition Triggers*. Ph.D. thesis, Eberhard Karls Universität Tübingen, Tübingen.

Tiemann, S., Kirsten, M., Beck, S., Hertrich, I., and Rolke, B. (2015). "Presupposition processing and accommodation: an experiment on *wieder* ('again') and consequences for other triggers," in *Experimental Perspectives on Presuppositions*, ed F. Schwarz (Cham; Heidelbergl New York, NY; Dordrecht; London: Springer), 39–65.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., et al. (2011). "Psycholinguistic evidence for presuppositions: on-line and offline data," in *Proceedings of Sinn & Bedeutung 15*, eds I. Reich, E. Horch, and D. Pauly (Saarbrücken: Saarland University Press), 581–597.

Tonhauser, J., Beaver, D., Roberts, C., and Simons, M. (2013). Towards a taxonomy of projective content. *Language* 89, 66–109. doi: 10.2307/23357722

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *J. Semantics* 9, 333–377.

Zeevat, H. (2002). "Explaining presupposition triggers," in *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, eds K. van Deemter and R. Kibble (Stanford, CA: CSLI Publications), 61–87.

Zehr, J., and Schwarz, F. (2016). "Entailed vs. non-entailed presuppositions: an experimental assessment," in *NELS 46: Proceedings of the Forty-Sixth Annual Meeting of the North East Linguistic Society, Vol. 3*, eds C. Hammerly and B. Prickett (Amherst, MA: GLSA), 319–328.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 May 2019
doi: 10.3389/fpsyg.2019.01227

Grasping the Alternative: Reaching and Eyegaze Reveal Children's Processing of Negation

*Alison W. Doyle, Kelsey Friesen, Sarah Reimer and Penny M. Pexman**

Department of Psychology, University of Calgary, Calgary, AB, Canada

There is evidence that children begin to understand negation early in the preschool years, but children's processing of negation is not well understood. We examined children's processing of denial negation using a variant of the visual world paradigm called the Shopping Task. In this task, participants help a puppet to find the items on a shopping list, selecting from two potential items on each trial in response to the puppet's affirmative ("the next item is an apple") or negation ("the next item is not an orange") sentence. In this binary decision context, participants' eye gaze and reaching behavior were tracked as they selected the item the puppet wants. Participants were 78 children aged 4–5 years and a comparison group of 30 adults. Results showed that children took longer to process negation than affirmative sentences, and that this difference arose early in processing. Further, children's eye gaze behavior suggested that on negation trials they regularly looked first to the negated object and were considering the negated meaning early in processing. Adults did not take longer to process negation than affirmative sentences, but their eye gaze behavior also indicated early consideration of negated meanings for negation sentences. We also examined relationships between children's language and executive function skills and their processing of negation and found no significant relationships. We conclude that both adults and children activate to-be-negated information in the processing of negation. Children, however, are less efficient at processing negation in this context.

Keywords: denial negation, negation processing, eye gaze, visual world, inhibitory control

OPEN ACCESS

Edited by:
Sophie Repp,
Universität zu Köln, Germany

Reviewed by:
Roman Feiman,
Brown University, United States
Chiara Gambi,
Cardiff University, United Kingdom

****Correspondence:***
Penny M. Pexman
pexman@ucalgary.ca

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *05 October 2018*
Accepted: *09 May 2019*
Published: *24 May 2019*

Citation:
Doyle AW, Friesen K, Reimer S and Pexman PM (2019) Grasping the Alternative: Reaching and Eyegaze Reveal Children's Processing of Negation. Front. Psychol. 10:1227. doi: 10.3389/fpsyg.2019.01227

INTRODUCTION

Negation is universal to human language, and is commonly used in both adult and child speech, usually in the form of "no," "not," and the suffix "-n't." Much of the developmental literature has focused on the production of negation (e.g., Klima and Bellugi, 1966; Pea, 1980; Bloom, 1993). This research has documented that many children begin using the word "no" as refusal or non-existence negation around 12 months of age (Dale and Fenson, 1996). Later, around 24–30 months of age, children begin producing denial negation (e.g., "this is not a puppy"), which can be more complex because it typically involves negation of something that was expected or at least plausible. Acquisition of other negative terms continues through the preschool years (e.g., Phillips and Pexman, 2015). Despite its importance to successful communication, there is much we do not yet know about how children come to understand negation. While children's production of negation is well documented, less research has explored children's processing of negation. In the present study we investigated children's processing of denial negation and the developmental skills that might support that processing. We focused specifically on contexts where there are only two (binary) interpretive possibilities for negation: the negated referent and the intended referent.

There is significant debate about the processing mechanism of negation comprehension. Much of the debate concerns whether, in understanding negation, the comprehender needs to activate the to-be-negated information. According to the multi-step processing account (e.g., Wason and Johnson-Laird, 1972; Carpenter and Just, 1975; Dudschig and Kaup, 2018), negation involves mentally representing the negated content or proposition (e.g., a closed cupboard for the sentence "The cupboard is not closed"), followed by its rejection, and finally representing the actual state of affairs (e.g., an open cupboard). By this position, negation will be more difficult to process than affirmative language, because it creates an information processing conflict (Dudschig and Kaup, 2018). In contrast, by the one-step account, it is not necessary that the to-be-negated information first be activated and then rejected when processing negation (e.g., Anderson et al., 2010; Tian et al., 2010; Papeo et al., 2016). The presence of negation markers may activate inhibitory processes that block activation of the negated referent and allow the actual state of affairs to be represented directly (Papeo et al., 2016). Some accounts suggest a third possibility, that the process normally involves multiple steps, except in cases where processing of negation is licensed pragmatically (e.g., Dale and Duran, 2011). In such cases, the negation concept and the target concept could be activated in parallel and eventually bound or fused into a representation for the true state of affairs (Anderson et al., 2010). By this account, one would expect the to-be-negated information to be activated in some contexts but less so in others.

Research on adults' processing of negation has provided evidence for multiple positions in this theoretical debate. For instance, studies show that adults are slower to respond to negation in picture/sentence verification tasks (Clark and Chase, 1972; Carpenter and Just, 1975). In these tasks, participants judge whether a sentence is a true or false description of a picture. Longer latencies for negated (e.g., "the star isn't above the plus") compared to affirmative statements (e.g., "the star is above the plus") in these tasks are taken as evidence that extra processing is required to understand negation, because the to-be-negated information must first be represented and then recoded. Thus, the findings were taken as evidence for a multi-step account.

The processing of negation was further explicated in a study by Kaup et al. (2006). Kaup et al. (2006) presented affirmative and negated sentences for self-paced reading. Each sentence was followed by a picture at one of two between-subjects delays: 750 or 1500 ms. Participants were asked to name the object in the picture, which either matched or mismatched the state of affairs described in the sentence. At the shorter 750 ms delay, following affirmative sentences (e.g., "the door is open"), participants were faster to name the pictures that matched the actual meaning (i.e., an open door) than the pictures that matched the alternate state (i.e., a closed door). No such difference was observed for negation sentences. However, at the longer 1500 ms delay, following negation sentences (e.g., "the door is not open"), participants were faster to name the pictures that matched the actual meaning (i.e., a closed door) than pictures that matched the negated state (i.e., an open door). The authors interpreted these effects to mean that because participants must first represent the negated state of affairs it takes longer for participants to focus attention on the correct meaning of negation sentences than affirmative sentences; eventually, participants are able to represent only the true state of affairs for negation statements.

In a recent study, Orenes et al. (2014) used a variant of the visual world paradigm (Huettig et al., 2011) with adult participants. Visual world tasks are characterized by the simultaneous presentation of verbal and visual stimuli. During stimulus presentation participants' eye movements are tracked. Research suggests that when verbal input is received, it is automatically processed and eye gaze shifts toward the visual referent. If negation is presented in the visual world paradigm, it is assumed that participants will tend to look to the referent that is most active at any given moment as they process the negated language. Using this paradigm, Orenes et al. (2014) showed that in binary context (when there are only two possible referents), participants initially fixated on the negated target and then shifted attention to the actual target. This shift in attention took time, and longer latencies were observed for negation than affirmative sentences. Similarly, in a subsequent eyetracking study Orenes et al. (2016) found that negation sentences were processed more slowly than affirmative sentences across several different pragmatic contexts. Thus, they concluded that the to-be-negated information must first be represented and rejected, so negation is always more difficult to process than affirmative language.

In contrast, other adult studies suggest that negation does not necessarily involve representation and rejection of the to-be-negated information and does not always take longer to process than affirmative language. Dale and Duran (2011) used mouse tracking to register responses in a true/false sentence verification task. When sentences were embedded in a context where participants had stronger expectations for negation, the mouse trajectories suggested that equivalent processes were involved for comprehension of negation and affirmative language, and negation did not take longer to process than affirmative language. In contrast, when the sentences were not embedded in strong context mouse trajectories showed evidence for shifts in interpretation during the response process and participants took longer to process negation than affirmative language. Dale and Duran inferred that negation *can* be processed readily when it is licensed pragmatically by context-based expectancies (for ERP data pointing to similar conclusions, see Nieuwland and Kuperberg, 2008). Similarly, Anderson et al. (2010) argued that the actual meaning of negation can be considered from the earliest moments of processing, without necessitating an initial stage where only the to-be-negated information is considered, especially if the negation is used in a situation where there are only binary alternatives.

Thus far, only a handful of studies have examined children's processing of negation. Kim (1985) showed that 3- to 5-year-old children took longer to process truth-functional negation (e.g., "this is not a car," presented with a picture of a ball) than true affirmative statements (e.g., "this is a ball"). Nordmeyer and Frank (2014) used eye tracking to examine children's comprehension of non-existence negation (e.g., "look at the boy with no apples"). Participants were 2- to 4-year-old children and a comparison group of adults. Only the 4-year-old children (and adults) showed good accuracy in their comprehension of negation. In terms of processing, children were less likely to look at the correct target for negated trials than for affirmative trials, but adults did not tend to show this difference. In interpreting their results, Nordmeyer and Frank noted that ". . . it is possible that when children are presented with two equally likely alternatives, identifying the referent of negation requires ruling out the named object" (p. 36). Thus, children may need to first consider, and rule out, the to-be-negated referent.

Different conclusions were drawn from the eyetracking study described by Reuter et al. (2018). Reuter et al. (2018) investigated processing of "didn't" by 2- and 3-year-old children. When Reuter et al. provided children with both pragmatic (e.g., story contexts that created expectations for negation) and semantic (blocking of affirmative and negated trials) supports, even 2-year-olds showed above chance accuracy interpreting negation. Further, children showed no differences in processing time for affirmative and negation sentences. Reuter et al. concluded that in pragmatically felicitous contexts it is not necessary for children to first process the to-be-negated meaning in order to understand negation.

Thus, the literature suggests that the processing of negation, for both adults and children, varies across different tasks and discourse contexts. One relevant factor seems to be the presence of context that supports negation. For instance, in adult research, tasks that provide pragmatic support for negation have tended to show equivalent processing times for negation and affirmative statements, leading to the inference that the meanings of negation can be considered directly. Although contextual factors likely explain some of the differences across studies, they do not account for all of the different patterns of results observed. Other relevant factors may include the type of negation, and how processing is measured. For instance, studies that examine only total processing time may miss processing differences between affirmative and negation sentences that could be revealed with other measures, like eye gaze.

Based on the developmental studies conducted thus far, it is not clear which theory is the best description of children's negation processing, how children's processing might be different from that of adults, and how children's processing of negation might be related to their language and cognitive skills. In the present research we approached these questions by investigating how 4- to 5-year-old children and a comparison group of adults process denial negation in a binary context. We used a variant of the visual world paradigm called the Shopping Task that we adapted from Kowatch et al. (2013). By this method, participants evaluate spoken sentences and select real objects based on their evaluations. In our version of the Shopping Task, participants listened to a puppet's directions about which of two objects the puppet wanted the participant to put in the shopping cart. Across trials, the puppet used both negation ("The next item is not candy") and affirmative ("The next item is carrots") sentences to indicate their wants. By age 4–5 years, we expected that children would have high levels of accuracy for comprehension of denial negation (Feiman et al., 2017), and that the focus of our investigation would be their response latencies and eye gaze fixations for correct responses. This method allowed us to measure the extent to which participants considered the to-be-negated meaning (the non-target object), as indicated by their looks to the non-target object for negation vs. affirmative sentences.

We also explored the linguistic and cognitive skills that might be related to children's processing of negation in order to better understand how negation processing develops. Children with more advanced language skills might be involved in more complex verbal interactions and thus develop more efficient negation processing. Further, inhibitory control involves the ability to reduce or override the influence of a non-target on active processes (for reviews see Diamond, 2013; Petersen et al., 2016) and is an important aspect of a child's executive functions, along with working memory and cognitive flexibility (Miyake et al., 2000). Such skills have been highlighted as factors that might support negation comprehension (e.g., Nordmeyer and Frank, 2014; Dudschig and Kaup, 2018) and these skills are developing rapidly in the age group we tested (e.g., Davidson et al., 2006; Zelazo and Carlson, 2012). In the present paradigm, inhibitory control and other executive function skills might help children to direct attention away from the negated object and toward the target object and thus may be related to their eye gaze on negation trials. To our knowledge, these hypothesized links between children's language, cognitive skills, and negation processing have not yet been tested.

MATERIALS AND METHODS

Participants

Participants were 78 children aged 4–5 years (M = 61.46 months, SD = 6.43) and a comparison group of 30 adults (M = 20.80 years, SD = 2.66). Children were recruited from the University of Calgary Child and Infant Learning and Development (ChILD) database. Children received two small toys as thanks for their participation. Adults were undergraduate students at the University of Calgary, recruited through the Psychology Department subject pool. Adults received bonus credit in a Psychology course in exchange for participation.

Procedure

Shopping Task

We measured participants' eye gaze and reaction time for negated and affirmative sentences in the Shopping Task. During the experiment participants were seated across a table from the experimenter. There were two toy food objects on the table, one placed on either side of a small toy shopping cart. The experimenter labeled the two food objects when they placed them on the table ("This time we have an apple and an orange"), thus establishing a pair of alternatives. Pete, a puppet worn on the experimenter's hand, was then introduced and participants were told that their task was to help Pete with his shopping:

Today you are going grocery shopping with Pete the Puppet. Your job is to listen to what Pete says and place the correct item into the basket. Each time Pete talks there will be an item placed on either side of the shopping cart. You'll need to listen carefully because sometimes Pete will communicate which item he wants by telling you the item he doesn't want, instead. In other words, sometimes Pete will say, "The next item is a coconut" and sometimes he will say "The next item is *not* a coconut."

With their hands resting on the table in "ready position," participants were instructed to listen and to select the object that corresponded to the content of the sentence. There were two practice trials followed by 12 experimental trials: 6 affirmative and 6 negation. Pete's voice was played on an audio track on a computer. Fifty-four versions of the stimulus materials were created, in order to present each food object as both the target and non-target for both affirmative and negation sentences across all participants and to vary correct target side and trial order.

During study development, a male Canadian English speaker who was naïve to the study purpose recorded the sentence stimuli in a quiet room. The speaker recorded two sentence stems: (1) "the next item is" and (2) "the next item is not," in order to ensure standardized sentence length up to the point where the food item was named in each sentence. The speaker then recorded the names of the 28 target food items (4 for the practice trials and 24 for the experiment). Using the program Audacity, the name of each food item was added to each of the sentence stems to create the negation and affirmative sentences. The food items were selected from cooking/kitchen toys in pairs that we judged to be of the same food type and of similar desirability for our child participants (e.g., apple and orange, candy and chocolate, see **Appendix** for full list).

Following the shopping task, three additional cognitive and language measures were presented to the child participants:

Red Dog/Blue Dog Stroop-Like Task

We adapted the Red Dog/Blue Dog Task from the "high inhibition" condition of a Stroop-like task described by Beveridge et al. (2002). A line drawing of a cartoon dog was printed in either red or blue fill, then copied onto 26 individual cards such that each card had one red or one blue colored dog on the front. The cards were laminated and stacked in a deck. The researcher held the deck of 26 red dog/blue dog cards, and presented each successive card upon hearing the participant's previous response, regardless of accuracy. No feedback was provided. The first two cards introduced the dogs: "My name is blue (/red)" was written above the image of the dog. Children were told that the blue-colored dog was called "Red" and the red-colored dog was called "Blue." Thus the task required that children inhibit the tendency to name the actual color of the dog, and instead use its (opposite) name. The participant was encouraged to say "Hi blue/red" to the first two dogs in order to practice using the name. For the following 24 cards, participants were asked to give the name of the dog upon presentation. The cards were arranged so that participants saw no more than two of the same color of dog in a row.

Dimensional Change Card Sort Task

The Dimensional Change Card Sort Task (DCCS) was administered as a measure of cognitive flexibility, following the instructions given in Zelazo (2006). Two small cardboard boxes, identical in size (15 cm × 15 cm × 15 cm) were used as card receptacles, each with a line drawing affixed to the front and back of the box. One box pictured a blue rabbit, the other a red boat. Cards were laminated and cut to 7 cm × 10.75 cm size. The researcher explained the rules of the pre-switch phase, counterbalanced (across participants) to start with either the "color game" or the "shape game." The researcher gave two example cards, labeling each one by its relevant dimension. For example, "This one is a red one, so it goes in the red box." Or, "this one is a rabbit, so it goes in the rabbit box." Participants were asked to sort each of the next six cards into their appropriate boxes, before the post-switch rules were given. To initiate the switch, participants were told "we aren't going to play the color (/shape) game any more, now we're going to play the shape (/color) game, where all the rabbits (/blue ones) go in this box, and all the boats (/red ones) go in this box." Six trials followed in the post-switch phase, and again the researcher labeled each card by its relevant dimension. For example, "here's a rabbit, where does it go?" All participants completed the pre- and post-switch phase and moved on to the border phase. One example of each type of card (border and no border) was shown to explain the border-phase rules, followed by 12 trials. The border rules were repeated before each of the 12 trials: "if there is a border, we play the color game, and if there isn't a border, we play the shape game."

Peabody Picture Vocabulary Test

The Peabody Picture Vocabulary Test (PPVT, 4th Edition) was administered as a measure of receptive vocabulary, as outlined in the test manual (Dunn and Dunn, 2007). Participants were familiarized with the procedure through two practice trials. Each test page in the flip book contained four pictures. Participants were shown each test page and when the experimenter named the target word participants were asked to point to the referent. Target words are grouped in sets of 12, and each set gets progressively more difficult. Participants began at set five to establish a basal set in which they made one error or fewer. If the basal set was not achieved at set five, the experimenter went down a set, until the basal set could be determined. The participant's ceiling set was identified once they made eight or more errors in one set.

Coding

Each trial was videorecorded via a digital video camera positioned behind the experimenter, with the participant's face and hands in view. Videos were coded frame-by-frame to assess participant accuracy, reaction time, and eye gaze. For reaction time, children's responses were divided into three phases corresponding to early, middle, and late processing (Climie and Pexman, 2008; Whalen et al., 2019): lift (onset of object name to initiation of lift), contact (lift to contact with object), and release (contact with object to its release). Thus the coding for reaction time began from the onset of the object name in the puppet's sentence and ended when the participant released the object into the shopping cart. This method allowed us to consider more than total response time for each trial; the initiation of a physical response at the end of the early phase does indicate that the participant has completed enough processing to begin to respond, but variability could still be observed in later phases of the response and if so we would interpret it as reflective of final verification processes. There were no trials in which a child made contact with the non-target and then altered their reach to grab the target, however, there were several instances in which children initially reached for one item and changed direction. In these cases, the reaction time phases were coded as usual. In addition, children's eye gaze fixations to the target, non-target, puppet, and extraneous objects were examined. For both affirmative and negation sentence types, eye gaze coding began at the onset of the object name and ended when the participant made physical contact with the object, signifying their choice (the early phases of processing). Eye gaze was coded for whether a participant looked at least once to each of the objects during the total eye gaze coding time and which item participants looked to first (target or non-target) after the object name onset. First look was coded using the method introduced in by Halberda (2006). More specifically, the coder moved forward one frame at a time in the videorecording from the onset of the object name until a participant looked at either the target or non-target (looks to other objects were ignored). Participants were most often looking at the puppets (60% of trials) before they heard the target. On an additional 20% of trials they were looking at the source of the audio (computer speaker) that played the narration. For the remaining 20% of trials, participants were looking at the target or non-target. Thus, for a small proportion of trials participants were already looking at one of the response objects at the onset of the "first look" recording window. We recorded their first look location here even if they were pre-fixated on a response object, since it was impossible in these cases to distinguish intentional from unintentional first looks. First looks to the target were coded as "1" while looks to the non-target were coded as "0". First look data for 10 child participants were unavailable due to a technical problem with the video files, thus first look was analyzed for 68 child participants (M = 60.91 months, SD = 6.75). Trials that were answered incorrectly (1.60%) were excluded from the eye gaze analyses.

A second coder evaluated 25% of the videorecordings (7.5 adult participants and 19.5 child participants). Interrater reliability was assessed using a two-way mixed-model consistency single-measures intraclass correlation coefficient (ICC; McGraw and Wong, 1996; Hallgren, 2012). The resulting ICC was in the excellent range for all variables evaluated (Cicchetti, 1994). For adults: likelihood of looking to target, ICC = 0.92; likelihood of looking to non-target, ICC = 0.91. For children: likelihood of looking to target, ICC = 0.95; likelihood of looking to non-target, ICC = 0.96.

RESULTS

All Shopping Task data were analyzed at the trial level using mixed effects regressions. Models were computed using the "lme4" package (Bates et al., 2015) in the statistical software R (R Core Team, 2017). All analyses used a maximal linear mixed effect model (Barr et al., 2013) and included random subject and item intercepts as well as by-subject and by-item random slopes for the effect of sentence type and age. We used mixed effects linear regression models to analyze the effect of sentence type (affirmative vs. negation), and for the analysis of children's data the models included children's age in months and the interaction of age and sentence type. This way, we could test whether statement type influenced children's reaching and looking behavior when age was also in the model. In separate regressions we examined whether children's processing of negation and, separately, affirmative language, was related to the three measures of children's cognitive/language development: Red/Dog Blue Dog accuracy, DCCS Border Phase accuracy, and PPVT raw score. Mean scores for all child participants on the cognitive and language measures are presented in **Table 1**. We used the "lmerTest" package (Kuznetsova et al., 2017) to generate p-values for models' fixed effects.

Adults

Reaction Times

We examined adults' total reaction times and also times for each of the three phases of the response (**Figure 1**). Results showed that for adults there was no difference in total reaction time for affirmative and negation sentences [β = 14.85 (SE = 40.90), t = 0.36, p = 0.782]. There were also no differences in reaction times for the early [β = 42.38 (SE = 43.86), t = 0.97, p = 0.335] and middle [β = 13.96 (SE = 16.46), t = 0.85, p = 0.402] phases of the response. For the late (release) phase, however, adults were faster for affirmative sentences than negation sentences [β = −41.66 (SE = 15.33), t = 2.72, p = 0.010].

Eye Gaze

Next, we examined adults' looking behavior during the early phase of processing. Our analyses focused on whether participants looked to the non-target, and which item (target or

TABLE 1 | Mean scores for child participants on the cognitive and language measures.

Measures	Mean	*SD*
Red Dog/Blue Dog Stroop-Like Task	15.88	8.04
Dimensional Change Card Sort Task – Shape	5.55	1.47
Dimensional Change Card Sort Task – Color	5.81	0.97
Dimensional Change Card Sort Task – Border	6.96	2.09
Peabody Picture Vocabulary Test (PPVT)	107.30	17.78

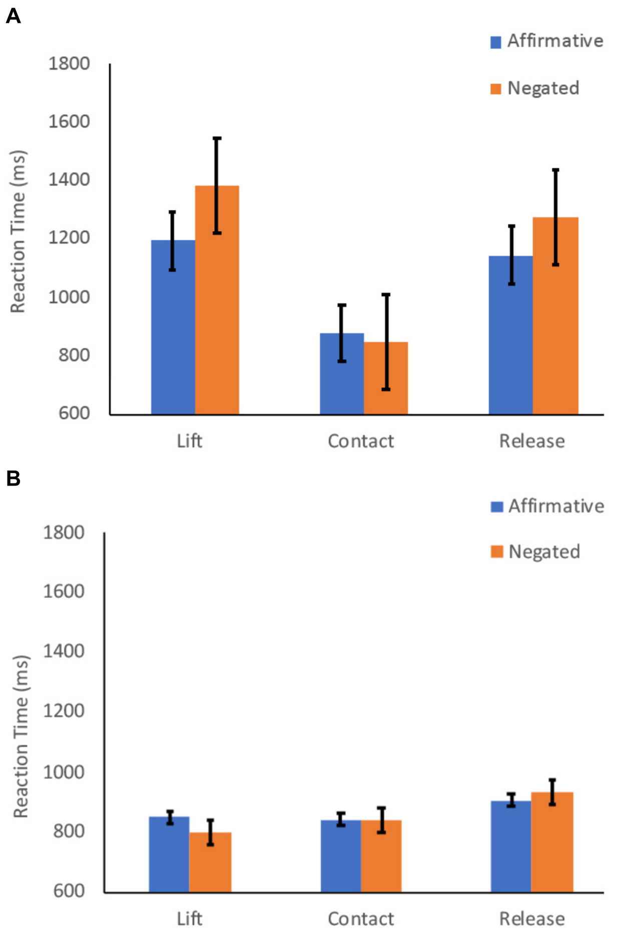


FIGURE 1 | Mean reaction times for **(A)** child and **(B)** adult participants for responses to affirmative and negated sentences in the Shopping Task. Reaching responses were divided into early (lift), middle (contact), and late (release) phases. Bars depict standard errors for mean reaction times in each phase.

non-target) they looked to first (see **Table 2** for mean likelihood of looking to each of the coded objects for negation and affirmative statements). Results from the logit regression showed that adults were more likely to look to the non-target during negation (M = 0.43, SD = 0.50) than affirmative (M = 0.28, SD = 0.45) sentences [β = −0.75 (SE = 0.24), z = −3.10, p = 0.002].

For first look, results showed that adults directed their first look to the non-target less than half of the time but did so more often during negation (M = 0.38, SD = 0.49) than affirmative (M = 0.19, SD = 0.39) sentences [β = 0.53 (SE = 0.13), z = 4.14, $p < 0.001$].

Children

Reaction Times

Analyses of children's total reaction times showed that children were faster to respond to affirmative than negation sentences [β = −129.99 (SE = 55.09), t = 2.36, p = 0.020]. The effect of children's age [β = −20.48 (SE = 13.68), t = 1.50, p = 0.14] was not significant for total reaction times, nor was the interaction of sentence type and age [β = 7.57 (SE = 6.45), t = 1.17, p = 0.241].

There was also a significant difference in reaction times for the early phase of processing of negation and affirmative sentences [β = −72.61 (SE = 35.55), t = 2.04, p = 0.043], such that early phase latencies were faster for the affirmative sentences than for the negation sentences. The effect of age was also significant [β = −17.56 (SE = 8.14), t = 2.16, p = 0.036] such that older age was associated with faster latencies in the early phase of processing. The interaction of sentence type and age [β = 4.07 (SE = 4.19), t = 0.97, p = 0.334] was not significant.

There was no significant difference in reaction times for affirmative and negation sentences in the middle [β = 11.86 (SE = 18.88), t = 0.63, p = 0.530] phase of children's responses. In this analysis, the effect of age [β = −5.92 (SE = 4.32), t = 1.37, p = 0.173] was also not significant nor was the interaction of age and sentence type [β = 0.11 (SE = 2.88), t = 0.04, p = 0.969]. There was no significant difference in reaction times for affirmative and negation sentences in the late [β = −57.03 (SE = 44.71), t = 1.28, p = 0.207] phase of children's responses. Again, neither the effect of age [β = −10.07 (SE = 7.83), t = 1.29, p = 0.201] nor the interaction of age and sentence type [β = 5.23 (SE = 5.94), t = 0.88, p = 0.381] were significant.

TABLE 2 | Mean likelihood of looking to each coded object for adult and child participants (standard deviations in parentheses).

	Adults		Children	
Objects	**Negation**	**Affirmative**	**Negation**	**Affirmative**
Target	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Non-target	0.43 (0.50)	0.28 (0.45)	0.71 (0.46)	0.58 (0.49)
Extraneous	0.64 (0.48)	0.60 (0.49)	0.94 (0.24)	0.92 (0.27)
Puppet	0.49 (0.50)	0.56 (0.50)	0.58 (0.49)	0.66 (0.47)

Eye Gaze

Next, we examined children's likelihood of looking to the non-target during the early phase of processing. Results of the logit regression showed that children were more likely to look to the non-target for negation (M = 0.71, SD = 0.46) than for affirmative (M = 0.58, SD = 0.49) sentences [β = −0.29 (SE = 0.08), z = −3.64, $p < 0.001$]. The effect of age was not significant [β = −0.012 (SE = 0.013), z = −0.91, p = 0.363], nor was the interaction of age and sentence type [β = 0.00 (SE = 0.011), z = 0.02, p = 0.983].

For first look, results showed that children directed more first looks to the non-target for negation (M = 0.53, SD = 0.50) than for affirmative (M = 0.36, SD = 0.48) sentences [β = 0.324 (SE = 0.08), z = −4.11, $p < 0.001$]. In this analysis, neither the effect of age [β = 0.002 (SE = 0.011), z = 0.25, p = 0.804] nor the interaction of age and sentence type [β = −0.013 (SE = 0.011), z = −1.22, p = 0.223] were significant.

Finally, we examined whether children's processing of negation and, separately, affirmative statements was related to their cognitive and language skills. The results are presented in **Table 3**. Given the number of analyses conducted here, once correction is applied for multiple comparisons, none of the results in **Table 3** could be considered significant.

TABLE 3 | LME model estimates for the effects of children's cognitive and language skills on processing of negation and affirmative sentences.

	Negation sentences			Affirmative sentences			Negation sentences			Affirmative sentences		
Fixed effect	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**
Total reaction time												
Age	−21.43	16.38	0.192	−25.37	24.03	0.295	–	–	–	–	–	–
PPVT	−10.88	8.25	0.192	−9.39	8.49	0.273	−14.33	7.97	0.077	−13.03	7.78	0.098
Red/Blue Dog	1.08	17.74	0.952	−15.83	17.62	0.372	−2.16	17.89	0.904	−18.60	17.43	0.290
DCCS Border	−50.71	68.19	0.460	51.67	68.27	0.452	−52.60	69.43	0.451	42.28	67.74	0.535
Lift reaction time												
Age	−3.38	6.59	0.567	−15.48	12.90	0.234	–	–	–	–	–	–
PPVT	−6.65	3.26	0.045	−5.34	4.56	0.246	−7.25	3.11	0.023	−7.56	4.19	0.075
Red/Blue Dog	−9.73	7.00	0.169	−16.61	9.47	0.084	−10.30	6.99	0.145	−18.29	9.39	0.055
DCCS Border	−15.66	26.91	0.563	35.12	36.67	0.342	−15.99	27.12	0.557	29.42	36.48	0.423
Contact reaction time												
Age	−5.61	5.42	0.302	−5.55	6.86	0.421	–	–	–	–	–	–
PPVT	0.74	2.33	0.750	0.35	2.43	0.886	−0.16	2.17	0.941	−0.44	2.21	0.842
Red/Blue Dog	0.95	4.91	0.847	2.85	5.02	0.573	0.122	4.88	0.980	2.23	4.95	0.653
DCCS Border	−22.31	18.81	0.239	−22.35	19.52	0.256	−22.77	18.91	0.233	−24.34	19.31	0.212
Release reaction time												
Age	−14.22	11.61	0.222	−4.14	12.36	0.739	–	–	–	–	–	–
PPVT	−4.51	5.37	0.403	−4.32	4.37	0.326	−6.80	5.08	0.185	−4.92	3.97	0.220
Red/Blue Dog	9.29	11.43	0.419	−2.48	9.05	0.785	7.14	11.42	0.534	−2.93	8.89	0.743
DCCS Border	−11.07	43.88	0.802	39.41	35.13	0.266	−12.23	44.31	0.783	37.89	34.62	0.277
Likelihood of looking to non-target												
Age	−0.0002	0.004	0.961	−0.0004	0.004	0.929	–	–	–	–	–	–
PPVT	−0.002	0.001	0.289	−0.002	0.001	0.246	−0.002	0.001	0.235	−0.002	0.001	0.189
Red/Blue Dog	−0.002	0.003	0.594	−0.006	0.003	0.082	−0.002	0.003	0.581	−0.006	0.003	0.075
DCCS Border	−0.02	0.01	0.071	−0.015	0.012	0.229	−0.023	0.012	0.070	−0.015	0.012	0.218
First look												
Age	0.002	0.004	0.603	−0.003	0.005	0.548	–	–	–	–	–	–
PPVT	0.002	0.002	0.307	0.0002	0.002	0.908	0.002	0.002	0.192	−0.0002	0.002	0.917
Red/Blue Dog	−0.002	0.004	0.642	−0.001	0.003	0.699	−0.001	0.003	0.697	−0.002	0.003	0.639
DCCS Border	0.026	0.014	0.068	0.012	0.014	0.264	0.026	0.014	0.066	0.014	0.013	0.295

In the leftmost six columns the analyses include age, whereas in the rightmost six columns the analyses do not include age.

DISCUSSION

The purpose of the present study was to examine children's processing of negation. In the Shopping Task, a pair of alternatives was established by the experimenter before the trial started, the speaker used negation on half of the trials, and expectations were created in the instructions that the speaker might communicate what they wanted by stating what they did *not* want. In addition, the task context was binary: on each trial, the speaker referred to one of only two objects.

Results showed that in this context there was evidence that children considered the to-be-negated information when processing negation. This was evident in the eye gaze data. Eye gaze analyses showed that children looked more often at both the target and the non-target for negation sentences than for affirmative sentences. In addition, children were slightly more likely to look first to the non-target than the target on negation trials (53%), and this tendency to look to the non-target first was significantly less common on affirmative trials. These eye gaze data suggest that children often processed the negated meaning before shifting gaze to the correct object. As such, the results could be taken as evidence for the possibility outlined by Nordmeyer and Frank (2014), mentioned here in the Introduction, that children presented with two viable alternatives will need to rule out the named object in order to correctly select the intended referent. In addition, we found that children took longer to process negation than affirmative language. Insights from total processing time can be limited, however, and so further insight was provided by the early phase latencies as these give us clues about what children were considering during early processing. Children's longer latencies for negation sentences were driven by delays early in processing (the "lift" phase), presumably because it took additional time to activate and then rule out the to-be-negated meaning. Together, these findings suggest that children's processing was best described by a multi-step account of negation processing.

Adults, too, showed evidence in their eye gaze that they often considered the to-be-negated meaning on negation trials, although perhaps less frequently than did children. Nonetheless, adults still occasionally looked first to the non-target (named object) on negation trials, and did so more often for negation than for affirmative trials. In addition, adults did not take longer to process negation overall, or in the early phases of processing. This suggests that adults were better able to deal with response conflict on negation trials than were children. In contrast, adults showed longer latencies for negation only in the final phase of processing ("release"). This could be taken as evidence for a final integration (fusing) or verification stage that is more time-consuming for negation sentences. As such, on balance, the adults' data could also be interpreted as consistent with a multi-step account of negation processing. The adult data are also consistent with the notion that there are circumstances where negation can be processed as quickly as affirmative language, such as when it is licensed pragmatically (e.g., Dale and Duran, 2011).

We also explored relationships between children's cognitive and language skills and their processing of negation, in what we believe is the first examination of this issue. We found no evidence for significant relationships between the measures of children's inhibitory control or cognitive flexibility and children's processing of negation. As such, we found no support for the hypothesis that stronger executive function skills might help children to direct attention away from the negated object and toward the target object and thus that these skills would be related to their eye gaze on negation trials. These null findings could be a function of limitations in the present study (e.g., measures chosen), but they do suggest that in the age range tested here other factors may be worth considering in terms of relationships with children's processing of negation. In addition, it is possible that if younger children were tested then the expected relationships between executive function skills and negation processing might be observed. Younger children would likely find both the Shopping Task and the cognitive assessments to be more challenging and thus their negation processing performance might be more sensitive to individual differences in executive function skills.

There is an extensive literature that has considered the role of alternatives (not just those involved in negation) in language processing. In negation, the use of the word *not* signals to the listener that alternatives should be activated. Given the present task context, there are only two alternatives (the target and the non-target) on each trial and participants likely pre-activate these when they are labeled by the experimenter at the trial's start. Thus, with the present form of negation there is probably less need for the kind of selection mechanisms that have been described in some of the other work on alternatives, where the listener needs to focus on contextually relevant alternatives, forming an alternative set (e.g., Husband and Ferreira, 2016; Gotzner and Spalek, 2017). The process of considering activated alternatives and suppressing the irrelevant meaning, however, would likely be similar for negation and other types of alternative resolution. One potential difference for negation was identified by Dennison and Schafer (2017). These authors compared adults' processing of intonationally implicated contrast (e.g., "The mailbox WAS full") with that of negation (e.g., "The mailbox was not full"). Results showed that the processing time course differed for the two statement types, with negation showing earlier differences in activation of the negated and correct meanings, and contrastive statements showing this difference later in processing. Dennison and Shafer speculated that this could be because the negated meaning, once rejected, does not need to be maintained for understanding ongoing discourse, whereas for contrastive statements the correct and negated meanings both have some relevance for understanding the ongoing discourse.

The results of the present study showed that both children and adults considered the target and non-target meanings early in processing of negation. Adults did not take longer to process negation than affirmative language. Children, however, did take longer to process negation than affirmative language. As such, we infer that while children in the present study were highly accurate at comprehending negation, their processing of negation was not yet as efficient as that of adults. It is possible that adults were better able to make use of the task context, as it licensed negation with speaker knowledge (the speaker description mentioned a tendency to use negation) and a high proportion of negated trials. In future research it will be important to identify the factors that contribute to children's developing efficiency in processing of negation, and to their emerging ability to draw inferences and derive expectations from the context in which language is used.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, with written informed consent from all adult participants and from caregivers on behalf of child participants. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Calgary Conjoint Faculties Research Ethics Board.

AUTHOR CONTRIBUTIONS

AD and PP conceived the study. AD, KF, and SR collected the data. AD, KF, and SR analyzed the data. AD, KF, SR, and PP wrote the manuscript.

FUNDING

This work was supported by a grant from the Social Sciences and Humanities Research Council (SSHRC) of Canada to PP grant no. 435-2013-0096.

ACKNOWLEDGMENTS

We acknowledge David Sidhu and Kristen Deschamps for assistance with stimulus preparation and coding, and the families who participated in this research.

REFERENCES

Anderson, S., Huette, S., Matlock, T., and Spivey, M. (2010). "On the temporal dynamics of negated perceptual simulations," in *Meaning form and Body*, eds F. Parrill, V. Tobin, and M. Turner (Stanford, CA: Stanford University), 1–20.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.

Beveridge, M., Jarrold, C., and Pettit, E. (2002). An experimental approach to executive fingerprinting in young children. *Infant Child Dev.* 11, 107–123. doi: 10.1002/icd.300

Bloom, L. (1993). *Language Development from Two to Three*. New York, NY: Cambridge University Press.

Carpenter, P. A., and Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychol. Rev.* 82, 45–73. doi: 10.1037/h0076248

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284–290. doi: 10.1037//1040-3590.6.4.284

Clark, H. H., and Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cogn. Psychol.* 3, 472–517. doi: 10.1016/0010-0285(72)90019-9

Climie, E. A., and Pexman, P. M. (2008). Eye gaze provides a window on children's understanding of verbal irony. *J. Cogn. Dev.* 9, 257–285. doi: 10.1080/15248370802247939

Dale, P. S., and Fenson, L. (1996). Lexical development norms for young children. *Behav. Res. Methods Instrum. Comput.* 28, 125–127. doi: 10.3758/bf03203646

Dale, R., and Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cogn. Sci.* 35, 983–996. doi: 10.1111/j.1551-6709.2010.01164.x

Davidson, M. C., Amso, D., Anderson, L. C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4-13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44, 2037–2078. doi: 10.1016/j.neuropsychologia.2006.02.006

Dennison, H., and Schafer, A. J. (2017). Processing intonationally implicated contrast versus negation in American English. *Lang. Speech* 60, 174–199. doi: 10.1177/0023830917694066

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Dudschig, C., and Kaup, B. (2018). How does "not left" become "right"? Electrophysiological evidence for a dynamic conflict-bound negation processing account. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 716–728. doi: 10.1037/xhp0000481

Dunn, L. M., and Dunn, D. M. (2007). *The Peabody Picture Vocabulary Test*, 4th Edn. Bloomington, MN: NCS Pearson, Inc.

Feiman, R., Mody, S., Sanborn, S., and Carey, S. (2017). What do you mean, No? Toddlers' comprehension of logical "No" and "Not". *Lang. Learn. Dev.* 13, 430–450. doi: 10.1080/15475441.2017.1317253

Gotzner, N., and Spalek, K. (2017). Role of contrastive and noncontrastive associates in the interpretation of focus particles. *Discourse Process.* 54, 638–654. doi: 10.1080/0163853x.2016.1148981

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cogn. Psychol.* 53, 310–344. doi: 10.1016/j.cogpsych.2006.04.003

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor. Quant. Methods Psychol.* 8, 23–34. doi: 10.20982/tqmp.08.1.p023

Huettig, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003

Husband, E. M., and Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Lang. Cogn. Neurosci.* 31, 217–235. doi: 10.1097/WCO.0b013e328316e6e0

Kaup, B., Lüdtke, J., and Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: is a door that is not open mentally closed? *J. Pragmat.* 38, 1033–1050. doi: 10.1016/j.pragma.2005.09.012

Kim, K. J. (1985). Development of the concept of truth-functional negation. *Dev. Psychol.* 21, 462–472. doi: 10.1371/journal.pone.0140658

Klima, E. S., and Bellugi, U. (1966). "Syntactic regularities in the speech of chidren," in *Psycholinguistic Papers*, Vol. 3765, eds J. Lyons and R. Wales (Edinburgh: Edinburgh University Press), 183–208.

Kowatch, K., Whalen, J. M., and Pexman, P. M. (2013). Irony comprehension in action: a new test of processing for verbal irony. *Discourse Process.* 50, 301–315. doi: 10.1080/0163853x.2013.799934

Kuznetsova, A., Brockhoff, P. B., and Christiensen, R. H. B. (2017). lmerTest Package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26.

McGraw, K. O., and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46. doi: 10.1037//1082-989x.1.1.30

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wagar, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Nieuwland, M. S., and Kuperberg, G. R. (2008). When the truth isn't too hard to handle: an event-related potential study on the pragmatics of negation. *Psychol. Sci.* 19, 1213–1218. doi: 10.1111/j.1467-9280.2008.02226.x

Nordmeyer, A. E., and Frank, M. C. (2014). The role of context in young children's comprehension of negation. *J. Mem. Lang.* 77, 25–39. doi: 10.1016/j.jml.2014.08.002

Orenes, I., Beltrán, D., and Santamaría, C. (2014). How negation is understood: evidence from the visual world paradigm. *J. Mem. Lang.* 74, 36–45. doi: 10.1080/17470218.2015.1063675

Orenes, I., Moxey, L., Scheepers, C., and Santamaría, C. (2016). Negation in context: evidence from the visual world paradigm. *Q. J. Exp. Psychol.* 69, 1082–1092. doi: 10.1080/17470218.2015.1063675

Papeo, L., Hochmann, J.-R., and Battelli, L. (2016). The default computation of negated meanings. *J. Cogn. Neurosci.* 28, 1980–1986. doi: 10.1162/jocn_a_01016

Pea, R. (1980). "The development of negation in early child language," in *The Social Foundations of Language and Thought*, ed. D. Olson (New York, NY: Norton), 156–186.

Petersen, I. T., Hoyniak, C. P., McQuillan, M. E., Bates, J. E., and Staples, A. D. (2016). Measuring the development of inhibitory control: the challenge of heterotypic continuity. *Dev. Rev.* 40, 25–71. doi: 10.1016/j.dr.2016.02.001

Phillips, C. I., and Pexman, P. M. (2015). When do children understand "opposite"? *J. Speech Lang. Hear. Res.* 58, 1233–1244. doi: 10.1044/2015_JSLHR-L-14-0222

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Available at: https://www.R-project.org/

Reuter, T., Feiman, R., and Snedeker, J. (2018). Getting to No: pragmatic and semantic factors in two- and three-year-olds' understanding of negation. *Child Dev.* 89, e364–e381. doi: 10.1111/cdev.12858

Tian, Y., Breheny, R., and Ferguson, H. J. (2010). Why we simulate negated information: a dynamic pragmatic account. *Q. J. Exp. Psychol.* 63, 2305–2312. doi: 10.1080/17470218.2010.525712

Wason, P. C., and Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press.

Whalen, J., Doyle, A., and Pexman, P. M. (2019). Sarcasm between siblings: children's use of relationship information in processing ironic remarks. *J. Pragmat.* (in press). doi: 10.1016/j.pragma.2019.05.005

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nat. Protoc.* 1, 297–301. doi: 10.1038/nprot.2006.46

Zelazo, P. D., and Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: development and plasticity. *Child Dev. Perspect.* 6, 354–360.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Practice Trials

(1) Peaches and pineapple

(2) Tuna [in can] and tomato sauce [in can]

Experimental Trials

(1) Chocolate donut and strawberry donut

(2) [slice of] Cherry pie and [slice of] blueberry pie

(3) Apple juice [in box] and orange juice [in box]

(4) Red pepper and corn on the cob

(5) Candy and chocolate

(6) Vanilla ice cream [on cone] and strawberry ice cream [on cone]

(7) Apple and orange

(8) Peas and carrots

(9) Potato chips and popcorn

(10) Pear and banana

(11) Creamy soup [in can] and vegetable soup [in can]

(12) Watermelon and grapes

ORIGINAL RESEARCH
published: 14 June 2019
doi: 10.3389/fpsyg.2019.01172

The Comprehension of Counterfactual Conditionals: Evidence From Eye-Tracking in the Visual World Paradigm

Isabel Orenes[1], Juan A. García-Madruga[2], Isabel Gómez-Veiga[2], Orlando Espino[3] and Ruth M. J. Byrne[4]*

[1] Department of Basic Psychology I, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, [2] Department of Developmental and Educational Psychology (UNED), Madrid, Spain, [3] Department of Psychology, Universidad de La Laguna, Santa Cruz de Tenerife, Spain, [4] School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Dublin, Ireland

Three experiments tracked participants' eye-movements to examine the time course of comprehension of the dual meaning of counterfactuals, such as "if there had been oranges then there would have been pears." Participants listened to conditionals while looking at images in the visual world paradigm, including an image of oranges and pears that corresponds to the counterfactual's conjecture, and one of no oranges and no pears that corresponds to its presumed facts, to establish at what point in time they consider each one. The results revealed striking individual differences: some participants looked at the negative image and the affirmative one, and some only at the affirmative image. The first experiment showed that participants who looked at the negative image increased their fixation on it within half a second. The second experiment showed they do so even without explicit instructions, and the third showed they do so even for printed words.

Keywords: counterfactuals, conditionals, comprehension, visual-world-paradigm, reasoning

OPEN ACCESS

Edited by:
Katharina Spalek,
Humboldt University of Berlin,
Germany

Reviewed by:
Brian Leahy,
Harvard University, United States
Bob Van Tiel,
Free University of Brussels, Belgium

****Correspondence:***
Isabel Orenes
iorenes@psi.uned.es

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *18 November 2018*
Accepted: *03 May 2019*
Published: *14 June 2019*

Citation:
Orenes I, García-Madruga JA, Gómez-Veiga I, Espino O and Byrne RMJ (2019) The Comprehension of Counterfactual Conditionals: Evidence From Eye-Tracking in the Visual World Paradigm. Front. Psychol. 10:1172. doi: 10.3389/fpsyg.2019.01172

INTRODUCTION

People often create counterfactual alternatives to reality in their everyday thoughts when they think "if only. . ." or "what if. . ." and imagine how a situation could have turned out differently (e.g., Kahneman and Tversky, 1982; Byrne, 2005). When people understand a counterfactual, such as "if there had been oranges then there would have been pears," they appear to envisage two possibilities, the imagined alternative to reality that corresponds to the counterfactual's conjecture, "there were oranges and pears" and the known or presumed facts that correspond to actual reality, "there were no oranges and no pears" (for a review see Byrne, 2016). In contrast, for a conditional in the indicative mood, such as "if there were oranges, then there were pears," they tend to envisage just a single possibility at the outset, "there were oranges and pears" (e.g., Johnson-Laird and Byrne, 2002; Khemlani et al., 2018). Our aim is to examine the mental representations and cognitive processes that underpin the comprehension of counterfactuals.

Our starting point is the extensive evidence for the dual meaning of counterfactuals. To fully understand the meaning of a counterfactual, people must not only simulate the imagined alternative to reality that is conjectured in a counterfactual, they must also recover the presumed reality.

But little is known about the cognitive processes that they rely on to do so (e.g., Johnson-Laird and Byrne, 1991; Byrne, 2005; Espino and Byrne, 2018). The accessibility of an imagined "possible world" from a representation of the real world poses difficulties (e.g., Stalnaker, 1968; Lewis, 1973), and what constitutes a "minimal change" is a slippery notion (e.g., Williamson, 2007; Kratzer, 2012). Nonetheless, people appear to readily recover the known or presumed facts when they understand a counterfactual. For example, participants tend to misremember a counterfactual, "if there had been oranges, then there would have been pears" and believe they were told instead, "there were no oranges and no pears" (Fillenbaum, 1974). They believe that someone uttering the counterfactual meant to imply this situation, and they judge that the items that best fit the description include this situation (e.g., Byrne and Tasso, 1999; Thompson and Byrne, 2002). Hence, the evidence indicates that they envisage the known or presumed facts, relying on their knowledge or on the cues of the subjunctive mood or content to do so.

People make more inferences that require access to "there were no oranges and no pears" from the counterfactual compared to the factual indicative conditional, such as *modus tollens* (from "there were no pears" to "therefore there were no oranges"). But they also make the same frequency of inferences that require access to "there were oranges and pears" from both conditionals, such as *modus ponens* – from "there were oranges" to "therefore there were pears" (e.g., Byrne and Tasso, 1999; Thompson and Byrne, 2002; see also Moreno-Ríos et al., 2008; Egan et al., 2009). They do so for various different sorts of content (e.g., Frosch and Byrne, 2012; see also Quelhas and Byrne, 2003; Egan and Byrne, 2012). Moreover, participants are primed to read, "there were no oranges and no pears" when they have first read the counterfactual, and they do so more quickly than when they have first read the factual conditional. But they also read, "there were oranges and pears" equally quickly from both sorts of conditional (e.g., Santamaría et al., 2005, see also Gómez-Veiga et al., 2010). Hence, the evidence indicates that people envisage both the imagined alternative to reality conjectured by the counterfactual, "there were oranges and there were pears," and the actual reality, known or presumed by the counterfactual, "there were no oranges and there were no pears." They keep track of their epistemic status as corresponding to real or imagined situations (Johnson-Laird and Byrne, 1991). The essence of the dual meaning of a counterfactual lies in this comparison of reality to an imagined alternative (e.g., Beck et al., 2006; Espino and Byrne, 2018). The question we wish to address is, at what point in their comprehension of a counterfactual do people detect the two messages of a counterfactual, that is, at what point do they envisage the conjecture, and at what point do they recover the presumed facts? The three experiments we report aim to advance knowledge of the comprehension of counterfactuals by establishing the point at which participants envisage each of the possibilities, during the temporal course of processing a counterfactual.

The question of when people envisage the situation corresponding to a counterfactual's conjecture and when they envisage the presumed facts is important, first because some theories dispute that people consider both possibilities, and second, because the time at which people consider each possibility can provide a clue about the cognitive processes that they rely on to do so. Some online comprehension studies have been interpreted to support the idea that people represent both the conjecture and the presumed facts, and others have been interpreted to suggest that they represent only the conjecture. On the one hand, in eye-tracking studies it has been found that participants looked at a target word more quickly when it was presented in a context that was consistent with the real world rather than a counterfactual world. The result indicates an early and fleeting reading-time penalty that appears to reflect the construction of two representations (e.g., Ferguson and Sanford, 2008), although it may be sensitive to methodological factors (e.g., Ferguson et al., 2008, 2010; Ferguson, 2012; see also Stewart et al., 2009). Similarly, counterfactuals elicit greater brain activation, compared to factual conditionals, in areas related to conflict detection (e.g., Kulakova et al., 2013; see also Urrutia et al., 2012b). The results suggest that people represent both possibilities. But on the other hand, false counterfactuals elicit more brain activity than true ones, which may indicate the activation of only the conjecture (e.g., Nieuwland and Martin, 2012; see also Nieuwland, 2013; Kulakova and Nieuwland, 2016a,b). Accordingly, some theorists have proposed that people understand a counterfactual by considering their belief only in the imagined alternative to reality and they do so by simulating only the conjecture (e.g., Evans and Over, 2004; Evans, 2007). Hence, one view is that only the conjecture about an imagined alternative to reality is represented; another view is that both the conjecture and the presumed facts of reality are represented.

Even among theorists who propose that people consider both possibilities, there are disagreements. One theory is that the counterfactual conjecture, "there were oranges and pears" is more highly activated than the presumed facts, "there were no oranges and no pears" (e.g., Ferguson, 2012; Ferguson and Cane, 2015). An alternative theory is that when people understand a counterfactual, they first represent the presumed facts, "there were no oranges and no pears," and the conjecture "there were oranges and pears," although activated, does not contribute to discourse updating, is not semantically integrated, and does not remain in focus following a delay (e.g., De Vega et al., 2007; De Vega and Urrutia, 2012; Urrutia et al., 2012a). Hence, one view is that the conjecture is the more highly activated of the two possibilities, whereas another view is that the presumed facts are more highly activated. Our aim is to contribute to the resolution of these conflicting ideas.

We address a novel and nuanced question in our three experiments: if people envisage both the counterfactual's conjecture about an imagined alternative to reality and its presumed facts, when do they do so? Our question is, at what point in the temporal process of comprehension do people consider each possibility? An answer to this question has the potential not only to distinguish between alternative theories of the comprehension of counterfactuals but also to shed light on the, at times, conflicting results of previous experiments, which have not been uniform in their choice of times at which to measure comprehension. Most importantly an answer to this question can contribute to understanding the nature of the cognitive processes by which people recover the presumed reality when they entertain the imagined alternative to reality conjectured by the counterfactual.

The three experiments we report rely on eye-tracking in the visual world paradigm to attempt to establish the time course during which people construct each possibility to understand a counterfactual. The visual world paradigm allows us to study the unfolding process of comprehending a counterfactual. In a typical visual world task, verbal and visual inputs are presented simultaneously while the participants' eye movements are recorded to provide an index of real-time processing, sensitive to subtle aspects of language, attention, and memory (e.g., Allopenna et al., 1998; Rayner, 1998; Duñabeitia et al., 2009; Orenes et al., 2014, 2015). The logic of the method is that when something is heard, it is processed and attended to automatically; at the same time, if the corresponding object is visible, the eyes begin to move toward it (e.g., Cooper, 1974; Tanenhaus et al., 1995; for a review see Huettig et al., 2011a). It follows that when people listen to counterfactuals in a visual world paradigm, they will look more frequently at the most active information in working memory.

In the three experiments we report, participants heard short stories that contained an indicative factual conditional, e.g., "if there are oranges, then there are pears," or a counterfactual, e.g., "if there had been oranges, then there would have been pears" (and we used a wide range of different contents, see the **Supplementary Material A**). In Experiment 1 and 2, four visual images related to the conditionals were shown on the screen, i.e., an image corresponding to the counterfactual's conjecture, e.g., of oranges and pears, and an image corresponding to the presumed facts, e.g., of no oranges and no pears, as well as "distractor" images, i.e., an image corresponding to other sorts of fruit, such as apples and strawberries, and an image corresponding to no apples and no strawberries, as **Figure 1** illustrates. The first two experiments differed in the instructions participants were given. In Experiment 1 participants were explicitly instructed to look at the objects on the screen that matched the meaning of the stories that they heard, in line with typical tasks in eye-tracking experiments, which encourage controlled, top-down processing. In Experiment 2 participants did not receive explicit instructions to look at the objects that matched the meaning of the stories, so that we could examine the processes underlying spontaneous counterfactual processing. In both experiments, we hypothesized that participants would look at the affirmative image, e.g., of oranges and pears, for indicative conditionals, whereas they would look at both the affirmative image and the negative image, e.g., of no oranges and no pears, for counterfactual conditionals. We anticipated that we would observe the same results with and without explicit instructions, notwithstanding an anticipated acceleration of processing of the counterfactual given explicit instructions. In Experiment 3, the same technique was used except that printed words were used instead of visual images, e.g., "oranges and pears." We again hypothesized that participants would look at the affirmative words for indicative conditionals, whereas they would look at both the affirmative and negative words for counterfactual conditionals. We anticipated that we would observe the same results for printed words, notwithstanding once again an anticipated acceleration of processing of the counterfactual given printed words, since images can impede the comprehension of negation (e.g., Orenes and Santamaría, 2014). Our key predictions concern the temporal course of looking at the affirmative and negative images as a participant hears a counterfactual. If the recovery of the presumed reality is essential to understanding the meaning of a counterfactual, then we expect to observe a rapid increase in looking at the negative image as soon as participants detect – for example through the cues of the subjunctive mood – that the conditional conveys an imagined alternative to reality. In addition, if the essence of the dual meaning of a counterfactual is the comparison of reality to an imagined alternative, then we expect to observe that participants will maintain their gaze on both the negative and the affirmative image throughout the period of time measured.

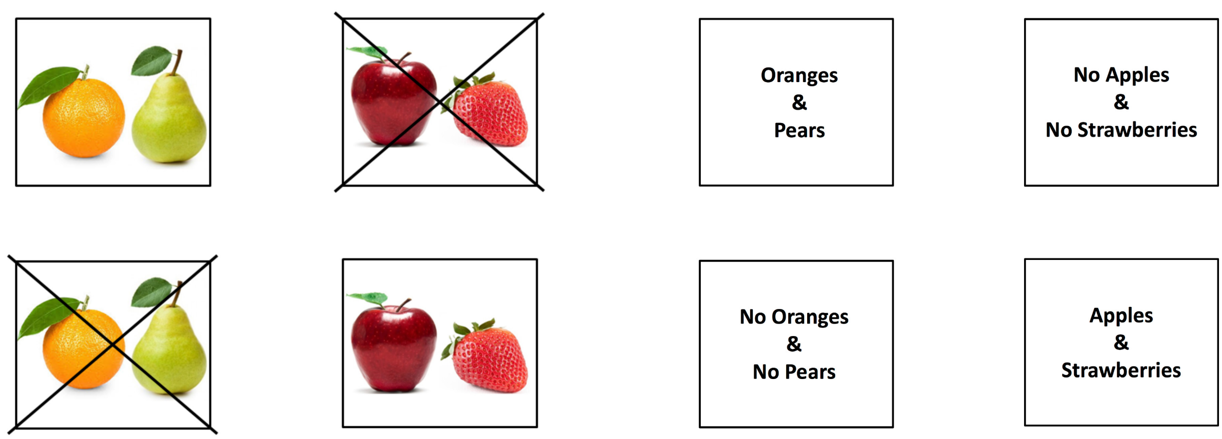


FIGURE 1 | Examples of the pictorial-based visual display used in Experiments 1 and 2 (on the left), and the word-based visual display used in Experiment 3 (on the right), for the counterfactual conditional "if there had been oranges then there would have been pears."

We analyzed participants' eye gaze at every 50 *ms* interval for a period of 4000 *ms*. One possibility is that a participant will look at only one image for a counterfactual during this time period, e.g., the affirmative image, and they will not look at the other three images. If so, the probabilities of fixations on the affirmative image will approximate 1, at every 50 *ms* "snapshot." We anticipate this outcome to be the case for indicative conditionals. An important possibility is that participants may look at two images, the affirmative one and the negative one. Even if they rapidly and constantly switch their eye gaze from the affirmative image to the negative image and back, our snapshot of fixations every 50 *ms* will capture their gaze on one image or the other at that precise time. Hence if participants look entirely equally at both images on every trial for each counterfactual, moving their gaze from one image to the other, the probabilities of fixations on the affirmative image and on the negative image will each approximate 0.5. We anticipate this outcome to be the case for counterfactual conditionals. Of course, the same is true if a participant looks only at the affirmative image for a counterfactual on one trial, and only at the negative image for a different counterfactual on another trial. Hence, we examine not only group data but also individual data in our experiments. If participants look entirely equally at all 4 images, including the distractors, the probabilities of fixations on each one of them will each approximate 0.25, although this outcome is unlikely given that listeners tend to fixate objects that are mentioned or expected.

EXPERIMENT 1

The aim of the experiment was to study the temporal course of the comprehension of counterfactuals. The question we wished to ask was, when people understand a counterfactual conditional, such as, "if there had been oranges then there would have been pears," at what point in the temporal course of processing do they focus on an image corresponding to the conjecture, e.g., "there were oranges and pears" and at what point do they focus on an image corresponding to the presumed facts, e.g., "there were no oranges and no pears." We expect that participants will begin by looking at the affirmative image for both counterfactual and indicative conditionals, since the items in the affirmative image match what is mentioned in the conditionals, but we predict that participants will exhibit a rapid increase in looking at the negative image as soon as they detect that the conditional conveys an imagined alternative to reality. We also predict that participants will continue to look at both the negative and the affirmative image throughout the measured period of time.

Methods

Participants

The participants were 24 volunteers who were students at the University of La Laguna, Tenerife, Spain, and they participated in the experiment in exchange for course credits. There were 21 women and 3 men, and their average age was 20 years, with a range from 18 to 26 years. The participants were native Spanish speakers and they all reported normal vision or wore soft contact lenses or glasses.

Materials and Design

The design was a within-participants one and participants received vignettes in each of two conditions: indicative or counterfactual conditionals. They heard 36 vignettes about simple events (adapted from Santamaría et al., 2005), 18 trials in each of the two conditions, and the order of the trials was randomized.

The vignettes were presented to the participants in their native Spanish and started with an opening scene, e.g., "Maria went to the fruit shop to buy fruit to make a cake for Valentine's Day. While she was waiting in the queue, she heard some clients who said" ("María fue a la frutería para comprar fruta que necesitaba para hacer un pastel por San Valentín. Mientras estaba esperando en la cola escuchó a unos clientes que decían"). The next sentence contained a conditional, either an indicative conditional, e.g., "if there are oranges, then there are pears" ("si hay naranjas, entonces hay peras") or a counterfactual conditional, e.g., "if there had been oranges, then there would have been pears" ("si hubiera habido naranjas, entonces habría habido peras"). The following sentence contained a conjunction, either an affirmative conjunction, e.g., "María realized that there were oranges and there were pears" ("María se dio cuenta que había naranjas y había peras") or a negative conjunction, e.g., "María realized that there were no oranges and there were no pears" ("María se dio cuenta que no había naranjas y no había peras"). The vignette ended with a closing-scene, e.g., "Finally, María also bought chocolate" ("Finalmente, María también compró chocolate"). The full set of 36 contents and their associated images is in the **Supplementary Material A**.

Each sentence was prerecorded and presented via speakers while four images were shown on a computer screen: two target images, e.g., an image of an orange and a pear, and an image of an orange and a pear with a cross through it, as well two distractor images, e.g., an image of other fruit such as an apple and a strawberry, and an image of an apple and a strawberry with a cross through it (see the four images on the left side of **Figure 1**). The position of each image (top left, top right, bottom left, bottom right quadrant) was counterbalanced across conditions. We constructed 8 versions of each vignette (e.g., oranges and pears) that varied in the conditional (indicative or counterfactual), the conjunction (affirmative or negative), and the reference to the objects in the set (e.g., oranges and pears, or apples and strawberries), as illustrated in **Table 1**. Each participant received only one of the 8 possible versions of each content in the set of 36 trials and the contents were assigned to the trials in a counterbalanced manner.

Apparatus and Procedure

Participants listened to the 36 stories over speakers while looking at a computer screen and at the end of each story they answered a simple question about it. Participants were instructed that they should listen to the sentences carefully and that they should not take their eyes off the screen throughout the experiment. They were explicitly instructed to look at the object or objects on the

TABLE 1 | Examples of the 8 versions of the verbal description of each content, illustrated for the oranges and pears/apples and strawberries content, for the visual display in **Figure 1**.

(1) Indicative affirmative	Oranges	If there are oranges, then there are pears. María realized that there were oranges and there were pears.
	Apples	If there are apples, then there are strawberries. María realized that there were apples and there were strawberries.
(2) Indicative negative	Oranges	If there are oranges, then there are pears. María realized that there were no oranges and there were no pears.
	Apples	If there are apples, then there are strawberries. María realized that there were no apples and there were no strawberries.
(3) Counterfactual affirmative	Oranges	If there had been oranges, then there would have been pears. María realized that there were oranges and there were pears.
	Apples	If there had been apples, then there would have been strawberries. María realized that there were apples and there were strawberries.
(4) Counterfactual negative	Oranges	If there had been oranges, then there would have been pears. María realized that there were no oranges and there were no pears.
	Apples	If there had been apples, then there would have been strawberries. María realized that there were no apples and there were no strawberries.

screen that matched the meaning of the stories that they heard. These explicit instructions were based on the usual information provided to participants in eye-tracking experiments, which typically specify how to interact with the display, e.g., by touching, clicking, or moving objects. Their eye movements were recorded at a rate of 500 Hz using an SR Research EyeLink II head-mounted eye-tracker connected to a 21 color CRT for visual stimulus presentation. Procedures were implemented in SR Research Experiment Builder. Calibration and validation procedures were carried out at the beginning of the experiment and were repeated several times per session. Trials started with the presentation of a central fixation dot for drift correction while participants listened to the opening scene sentence. Next, a display with four images appeared for 2 s. Then the story began, and the images remained on screen for the entire time while the remainder of the story was heard over speakers. The trial concluded with the appearance of a simple question on the screen, e.g., "Did María go to the fruit shop?" ("¿Fue María a la frutería?")[1], which participants answered by pressing either a "yes" or a "no" button on a game-pad, by pressing with their right index finger for yes and their left index finger for no. There was a practice block of four trials before the experiment proper started. The experiment lasted approximately 30–40 min and each participant was tested individually.

Results and Discussion

The data for the three experiments is available at https://reasoningandimagination.com/data-archive/ and on OSF at https://osf.io/n6hk3/. Prior to any data analysis one participant was eliminated from the analysis because her eye-movements explored the screen continuously without any systematic fixations on any point.

Eye-Tracking Data Coding

The eye-movement data generated by the EyeLink system were analyzed as follows. First, bitmap templates were created for identifying regions of interest in each display (the four pictures of the screen, e.g., oranges and pears, no oranges and no pears, apples and strawberries, and no apples and no strawberries). The object regions were defined in terms of rectangles containing the relevant objects, fixations landing within the perimeters of these rectangles were coded as fixations on the relevant objects. The output of the eye-tracker included the x- and y-coordinates of participant fixations, which were converted into region of interest codes using the templates.

The analysis of fixations was time-locked to the onset of the first object in the conditional, e.g., the onset of "oranges" in "if there *are/had been* oranges" and continued to 4000 *ms* after that word, which included listening to the rest of the conditional, "then there *are/would have been* pears" followed by a silent period. The periods were divided into 50 *ms* time slots and for each time slot, the number of fixations on each rectangle quadrant of the image was counted and converted into fixation probabilities[2].

To avoid problems inherent in proportional data, participant averages were arcsin-transformed prior to *t*-test comparisons. Given that 180–200 *ms* are usually assumed to account for saccade programming (Martin et al., 1993), the mean of the first time-region (0–100 *ms*) was considered to be the baseline and was used to conduct statistical comparisons against means on each time points at 50 *ms* intervals until 4000 *ms* later (for a similar method, see Huettig and Altmann, 2011). This correction to baseline allowed us to control for any bias in the pattern of fixations on images caused by the type of context. A false discovery rate (FDR) thresholding procedure, referred to as *pFDR-corr,* was used as an alpha correction to control for Type 1 errors due to multiple comparisons (81 for each condition; see Genovese et al., 2002).

T-Tests Against Baseline

The results reveal that participants looked at very different parts of the visual images on screen when they heard an indicative conditional compared to when they heard a counterfactual conditional, as **Figure 2** shows. For indicative conditionals, at the onset of the target word (e.g., oranges), participants were focused

[1]As a check that participants were attending to the task, we analyzed the response accuracy and latency for these simple comprehension questions in two repeated measures analysis of variance (ANOVA) of a 2 (conditional type: indicative vs counterfactual) × 2 (conjunction type: negative vs affirmative) design. They confirmed no effects of conditional type or conjunction type, nor any interaction between the two variables, $F<1$ in every case, and the results are presented in **Supplementary Table S1** in the **Supplementary Material B**.

[2]For the record, we also carried out an analysis of fixations time-locked to the onset of the affirmative or negative conjunction, e.g., the onset of the first instance of the word "there" in "there were (no) oranges and there were (no) pears" to 3000 *ms* after that word, which was the maximum duration of the conjunctions (the average was 1.6 *ms* for affirmative and 2.2 *ms* for negative conjunctions). The analysis is consistent with the main results and for brevity and completeness we report it in the **Supplementary Material B**.

on the affirmative image (oranges and pears) and the affirmative distractor (apples and strawberries) about equally frequently, with probabilities of fixation of about 0.4, as **Figure 2A** shows. This starting point may merely reflect a tendency to look at what is present rather than what is not present. What is revealing is that very early on in the process, 450 *ms* after the target word onset, the probabilities of fixation on the affirmative image started to increase (*pFDR-corr* = 0.002); fixations decreased on all other images, including the negative image (from 350 *ms*, *pFDR corr* = 0.034) (see the **Supplementary Material B** for details of the comparisons for the distractor images). The results show that for an ordinary indicative conditional, participants increase their fixation on the affirmative image very early indeed in the temporal course of processing, and fixations on the other three images decrease rapidly, as **Figure 2A** shows. Their fixation on the affirmative image continued throughout the period of measurement to 4000 *ms*.

A very different pattern emerges for counterfactual conditionals, e.g., "if there had been oranges, then there would have been pears." At the onset of the target word, e.g., "oranges," participants were focused on the affirmative image (oranges and pears) and the affirmative distractor (apples and strawberries) equally frequently with probabilities of fixations of about 0.4, as **Figure 2B** shows. Most revealingly, from very early on, 300 *ms* after the target word onset, the probabilities of fixation on the affirmative image started to increase (*pFDR-corr* = 0.039), and fixation on the negative image also *increased* (from 650 *ms*, *pFDR-corr* = 0.031). The results show that early in the temporal course of processing, within about half a second, participants increase their fixation not only on the affirmative image but also on the negative image; fixations on the two distractor images decrease rapidly, as **Figure 2B** shows (see the **Supplementary Material B** for details about the distractor images). Equally importantly, their

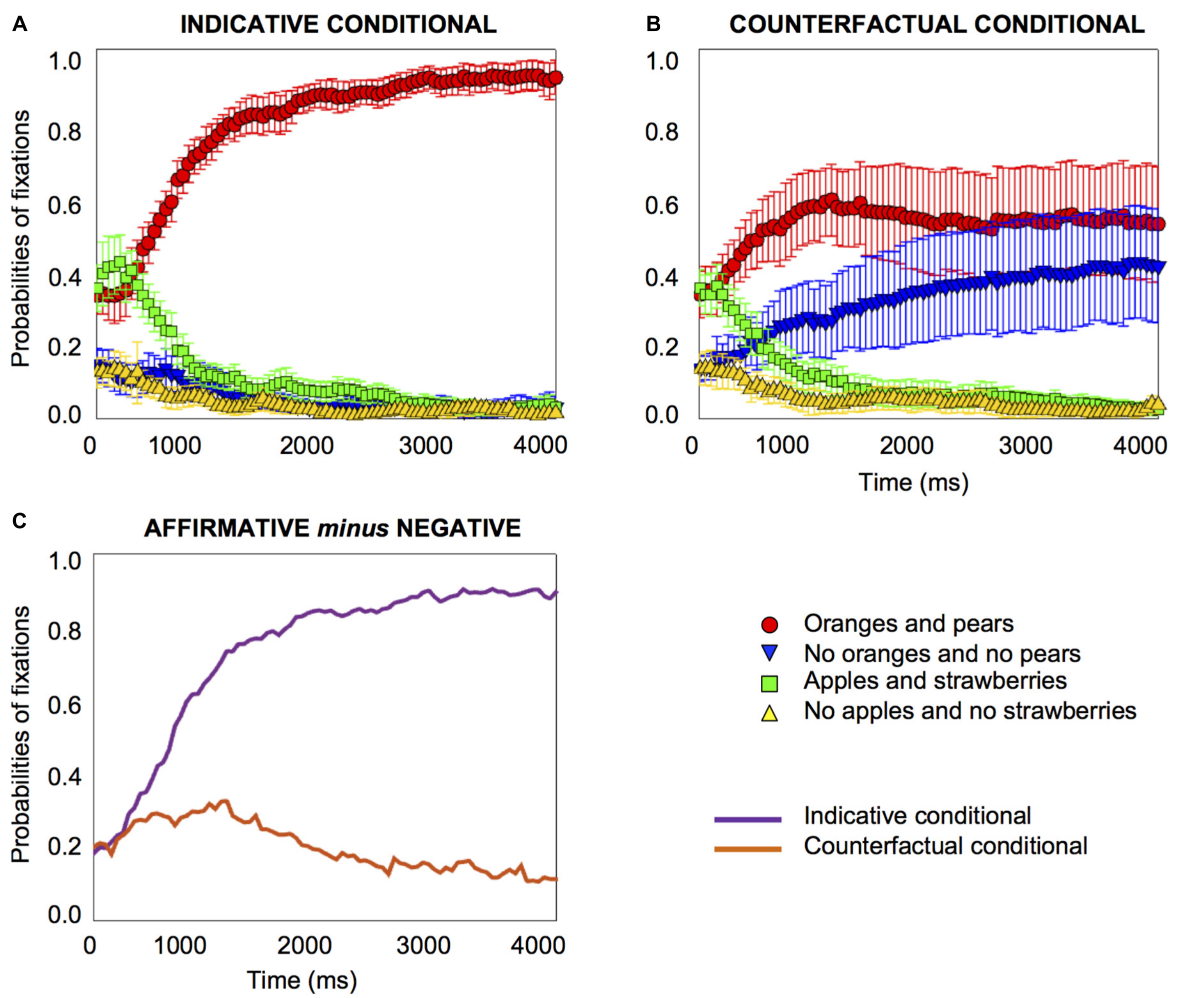


FIGURE 2 | Probabilities of fixations for indicative conditionals, e.g., "if there are oranges, then there are pears" **(A)** and counterfactual conditionals, e.g., "if there had been oranges, then there would have been pears" **(B)** in Experiment 1 on the affirmative and negative images, and the two distractor images, time-locked to the onset of the first object word, e.g., oranges. The differences in the probabilities of fixations (on the affirmative image minus the negative image) for indicative and counterfactual conditionals emerge at 850 *ms*, as **(C)** shows. Error bars are 95% confidence intervals within participants (see Morey, 2008; O'Brien and Cousineau, 2014).

fixation on both images continues throughout the period of measurement to 4000 *ms*.

Growth-Curve Analysis

We carried out a growth-curve analysis (Mirman, 2014; see the **Supplementary Material B** for details) which showed that people looked at the affirmative image more for the indicative conditional than the counterfactual, and they looked at the negative image more for the counterfactual than the indicative conditional. The increase of fixations on the affirmative image occurred more quickly for the indicative than the counterfactual conditional, and the opposite was the case for the negative image.

Analysis by Items

As a check that each of the 36 contents was interpreted in essentially the same way, we also carried out a similar analysis to compare indicative and counterfactual conditionals with *t*-tests against the baseline, but this time by items rather than by participants. It showed the same results. For indicative conditionals, at the onset of the target word (e.g., oranges), participants' focus was on the affirmative image and the affirmative distractor equally frequently. From 450 *ms* after the target word onset, the probabilities of fixation on the affirmative image started to increase (*pFDR-corr* = 0.003); fixations on the other images decreased, including for the negative image (from 1050 *ms*, *pFDR-corr* = 0.034) (see the **Supplementary Material B** for details about the distractor images).

For counterfactual conditionals, at the onset of the target word, the focus was on the affirmative image and the affirmative distractor equally frequently. From 300 *ms* after the target word onset, the probabilities of fixation on the affirmative image started to increase (*pFDR-corr* = 0.013) and so did fixation on the negative image (from 500 *ms*, *pFDR-corr* = 0.017), as **Figure 3** shows. It is noteworthy that the 95% confidence interval error is very much reduced for the counterfactuals in the by-item analysis compared to the by-participants one, as a glance at **Figures 2B**, **3B** shows, which suggests that the variance originates in differences between participants rather than differences between items. Accordingly, we also carried out an analysis of individual differences.

Individual Differences Analysis

We plotted individual graphs for each of the 23 participants, which are provided in the **Supplementary Material C**. As these graphs show, about half of the participants (n = 11) tended to look at the affirmative image only when they heard the counterfactual, just as they did for the indicative conditional; the other half of the participants (n = 12) tended to look at the negative image only (seven participants), or at both the affirmative and the negative image (five participants) for the counterfactual. We combined participants who looked at the negative image only and those who looked at the affirmative and negative image into a single sub-set group because consideration of the negative image (corresponding to the presumed facts of a counterfactual) indicates that individuals have reached a counterfactual interpretation of the conditional (and there are in any case too few participants to create three separate groups for reliable statistical analysis). These two sub-set groups of participants, affirmative only versus negative or negative-plus-affirmative, exhibited very different fixation patterns on the affirmative and negative image, as **Figure 4** shows.

Both groups showed similar patterns for the indicative conditional, the probabilities of fixation on the affirmative image started to increase early on (group 1 from 450 *ms*, *pFDR-corr* = 0.025; group 2 from 450 *ms*, *pFDR-corr* = 0.020), and to decrease on the other three images, including the negative image (group 1 from 800 *ms*, *pFDR-corr* = 0.037; group 2 from 1200 *ms*, *pFDR-corr* = 0.034) (see the **Supplementary Material B** for details about the distractor images). However, the two groups showed different patterns for the counterfactual. Group 1's pattern was the same as for the indicative: probabilities of fixation on the affirmative image started to increase early on (from 350 *ms*, *pFDR-corr* = 0.030), and probabilities of fixations on the other three images decreased, including for the negative image (from 1450 *ms*, *pFDR-corr* = 0.023). Group 2's pattern was different: probabilities of fixation for the affirmative image showed no significant changes to the baseline, but they *increased* for the negative image (from 400 *ms*, *pFDR-corr* = 0.039).

The analysis shows that when people understand the indicative conditional, they look at the affirmative image from very early (450 *ms*) and decrease their fixations on the negative image quite some time later (800 *ms* in Group 1; 1200 *ms* in Group 2). When they understand the counterfactual, one subset of participants do the same thing as for the indicative, they look at the affirmative image from very early (350 *ms*) and decrease their fixations on the negative image quite some time later (1450 *ms*); however, the other subset of participants look at the affirmative image early and continue to do so at the same rate as at the baseline throughout, but these participants look increasingly at the negative image and from very early indeed (400 *ms*).

The experiment provides information on the points in the temporal course of processing a counterfactual, e.g., "if there had been oranges, then there would have been pears," when people focus on an image corresponding to the conjecture, e.g., "there were oranges and pears" and on an image corresponding to the presumed facts, e.g., "there were no oranges and no pears." The results show that when people understand an indicative conditional, e.g., "if there are oranges, then there are pears," shortly after they hear the word "oranges," their focus increases on the affirmative image, and their focus on the other three images decreases fairly rapidly. The overall group results for indicative conditionals are reflected also in the results for each individual. The results show a different pattern for counterfactuals. The results averaged over the whole group of participants show that when they understand the counterfactual, e.g., "if there had been oranges, then there would have been pears," very early in the temporal course of processing, they increase their focus not only on the affirmative image but also on the negative image, and this focus on both images continues throughout the period of measurement. However, there are pronounced individual differences. About half of the participants appear to understand the counterfactual just as they do the indicative conditional, and they focus on the affirmative image only. The other half of the participants understand the counterfactual differently from the

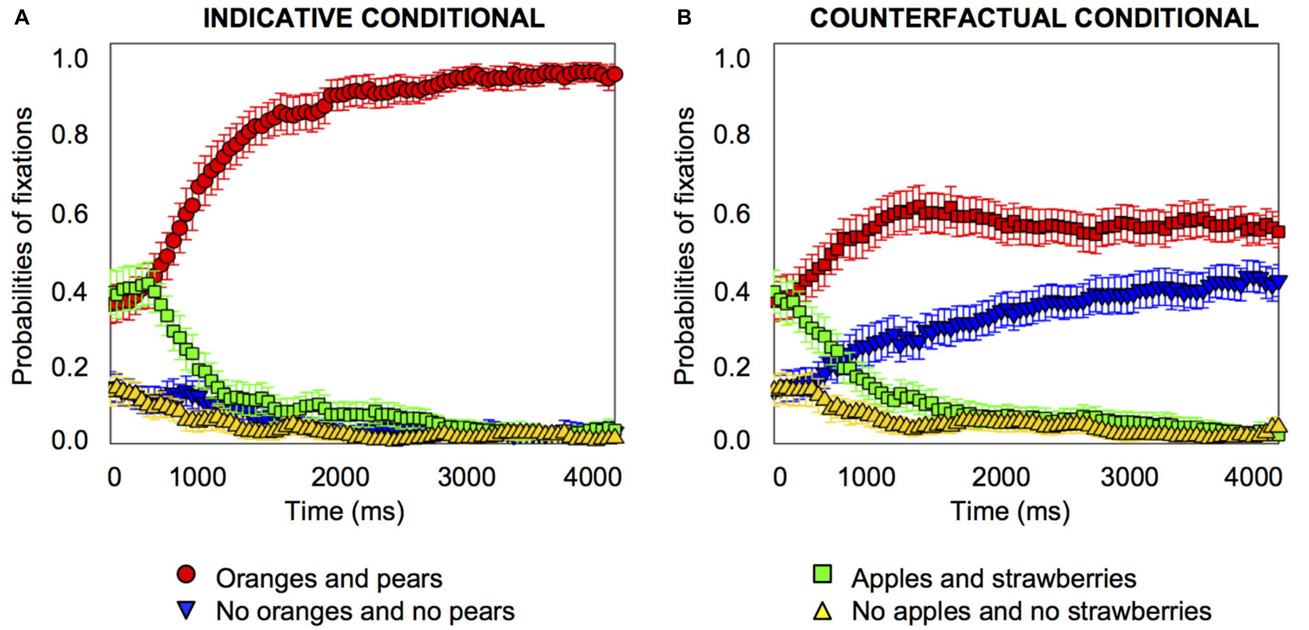


FIGURE 3 | Item-analysis probabilities of fixations for indicative conditionals, e.g., "if there are oranges, then there are pears" **(A)** and counterfactual conditionals, e.g., "if there had been oranges, then there would have been pears" **(B)** in Experiment 1. Error bars are 95% confidence intervals within items.

indicative conditional – they continue to look at the affirmative image as much as they do at the outset, but they increase their focus on the negative image.

One possible explanation for the individual differences is that the instruction to look at what the stories mean may be interpreted by some participants to look at what is explicitly mentioned in the counterfactual, e.g., oranges and pears, whereas it may be interpreted by others to look at what is presumed by the counterfactual, e.g., no oranges and no pears. To rule out this possibility, we carried out a second experiment with the aim of

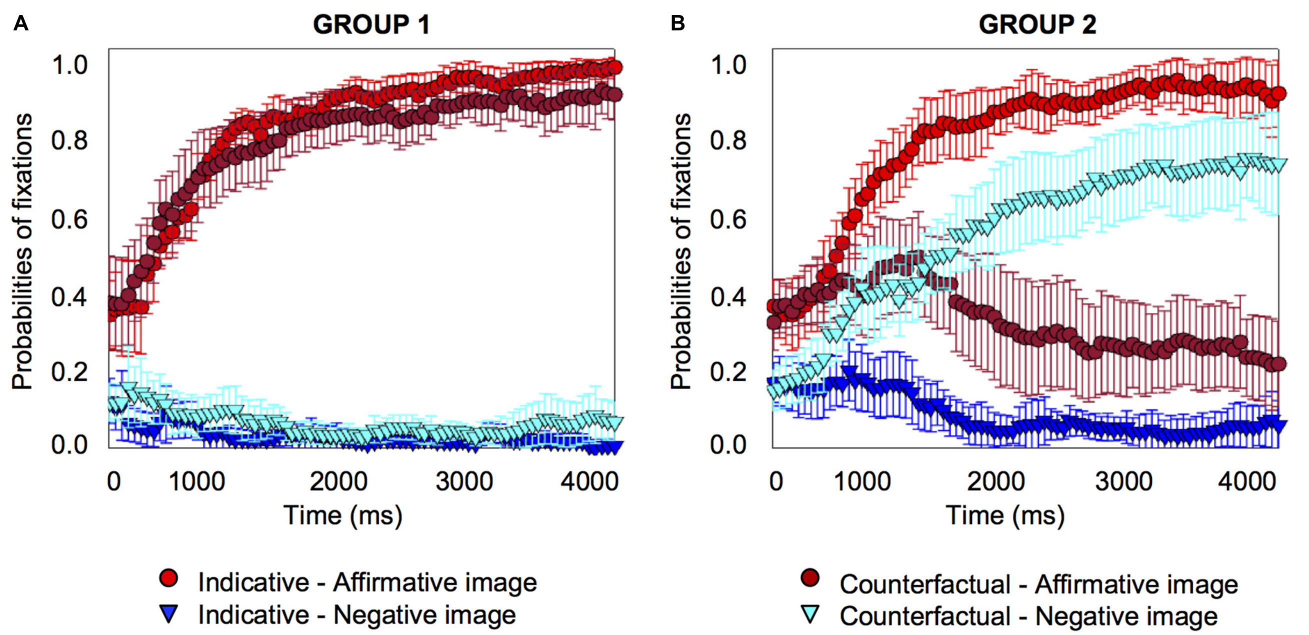


FIGURE 4 | Individual differences probabilities of fixations for indicative and counterfactual sentences for one subset group of 11 participants who looked at the affirmative image for counterfactuals **(A)**, and a second subset group of 12 participants who looked at the negative image only or at the affirmative and negative image for counterfactuals **(B)** in Experiment 1. Error bars are 95% confidence intervals within participants.

testing whether the results are replicated when participants are not given this explicit instruction.

EXPERIMENT 2

The aim of the experiment was to test whether the results of the previous experiment are replicated in an implicit task, that is, when participants are *not* given an explicit instruction to look at the object or objects on the screen that matched the meaning of the stories that they heard. In this way we aimed to examine further the spontaneous or automatic processes underlying the comprehension of counterfactuals.

Methods

Participants

The participants were a new set of 24 native Spanish speakers from the University of La Laguna, Tenerife, Spain, who participated in the experiment in exchange for course credits. There were 15 women and 9 men, and their average age was 19 years, with a range from 18 to 23 years. All of them reported normal vision or wore soft contact lenses or glasses.

Materials, Design and Procedure

The materials, design and procedure were the same as Experiment 1. The only difference was that participants were *not* instructed to look at the object or objects on the screen that matched the meaning of the stories they heard, as the participants in the previous experiment had been instructed. Participants were instructed to listen to the sentences and answer the simple question at the end. They were also told not to take their eyes off the screen throughout the experiment.

Results and Discussion

One participant was eliminated from the analysis because she looked at just one point on the screen throughout the experiment and no moves were registered for her, and five participants were eliminated because they explored the screen continuously without fixations on any point. The procedure for analyzing the eye movement data was the same as that used in the previous experiment.

T-Tests Against Baseline

The results replicated the previous experiment, as **Figure 5** shows. For the indicative conditional, at the onset of the target word, participants were focused on the affirmative image and the affirmative distractor equally frequently, with a probability of 0.3 to 0.4. From 400 *ms* after the target word onset, the probabilities of fixation on the affirmative image started to increase (*pFDR-corr* = 0.023); no significant change was observed for the negative image, as **Figure 5A** shows (see **Supplementary Material B** for details about the distractor images). Hence, participants looked at the affirmative image for the indicative conditional, replicating the findings of the previous experiment.

For the counterfactual, at the onset of the target word, participants were focused on the affirmative image and the affirmative distractor equally frequently, with a probability of 0.3 to 0.4. After the target word onset, there was an increase in fixations on the affirmative image (from 550 *ms*, *pFDR-corr* = 0.033), and an *increase* on the negative image (from 450 *ms*, *pFDR-corr* = 0.034). Hence, participants looked at the affirmative and the negative image for the counterfactual, replicating the findings of the previous experiment, as **Figure 5B** shows.

Growth-Curve Analysis

The growth curve analysis showed the same results as the previous experiment (see the **Supplementary Material B** for details). However, as **Figure 5C** shows, the differences between the indicative and counterfactual conditionals emerge at 1700 *ms*, which is considerably later than in the previous experiment (850 *ms*). Participants maintained their gaze on the affirmative image for both types of conditional until the negative image started to be fixated at a later time. This result reflects the difference in instructions between the two experiments and indicates that the instruction in the previous experiment to look at the objects that correspond to what the sentence means resulted in an earlier focus on the negative image in the understanding of the counterfactuals.

The results provide information on how people understand indicative and counterfactual conditionals and also reveal important clues about the implicit processes in the comprehension of counterfactuals, without explicit instruction. Despite the absence of instruction to look at the objects corresponding to what the sentence means, there is consistency in the results of this experiment and the previous one. The results suggest that people automatically look at the images that correspond to what they understand in this situation; when they are given instructions to do so explicitly, their processing of the sentences is accelerated, but the processing nonetheless remains the same.

Individual Differences Analysis

We again plotted individual graphs for each of the 18 participants, which are provided in the **Supplementary Material C**. Once again about half of the participants looked at the affirmative image only, when they heard the counterfactual (10 participants), just as they did for the indicative conditional; the other half looked at the negative image (eight participants, four who looked at the negative image and four who looked at both the negative and the affirmative image), as **Figure 6** shows.

For the indicative conditional, both groups showed the same pattern: the probabilities of fixation on the affirmative image increased (group 1 from 250 *ms*, *pFDR-corr* = 0.044; group 2 from 750 *ms*, *pFDR-corr* = 0.022) and fixations on the other images decreased, including for the negative image (group 1, no significant change; group 2 from 300 *ms*, *pFDR-corr* < 0.001) (see the **Supplementary Material B** for details about the distractor images). For the counterfactual, the groups showed different patterns. For group 1, the pattern was the same as the indicative conditional, the probabilities of fixation on the affirmative image increased (from 350 *ms, pFDR-corr* = 0.029), for the negative image there was no significant change. Group 2's pattern was different: probabilities of fixation for the affirmative image showed no significant changes to the baseline, but they *increased*

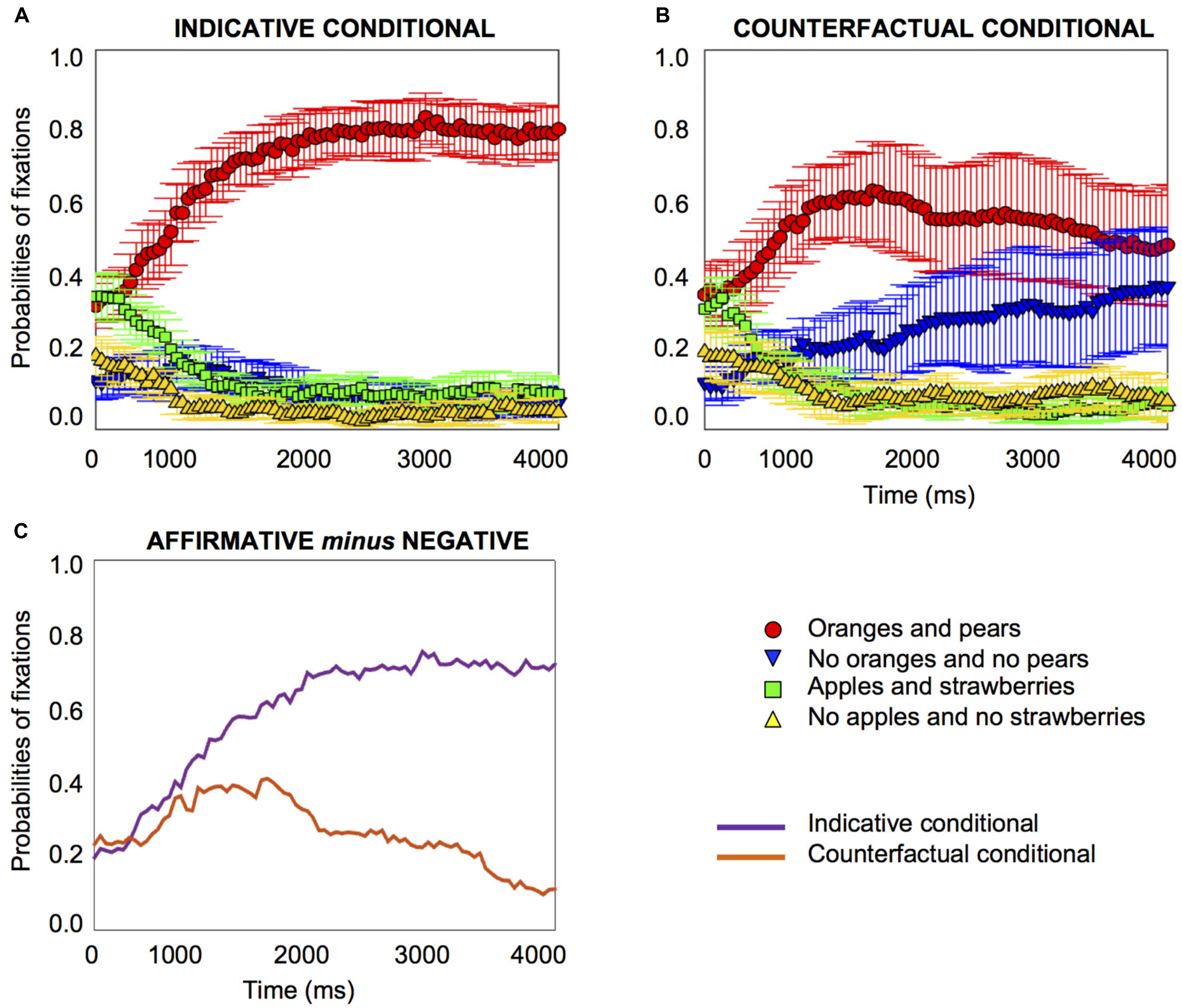


FIGURE 5 | Probabilities of fixations for indicative conditionals, e.g., "if there are oranges, then there are pears" **(A)** and counterfactual conditionals, e.g., "if there had been oranges, then there would have been pears" **(B)** in Experiment 2. The differences in the probabilities of fixations (on the affirmative image minus the negative image) for indicative and counterfactual conditionals emerge at 1700 ms, as **(C)** shows. Error bars are 95% confidence intervals within participants.

for the negative image (from 700 *ms, pFDR-corr* = 0.027). The results are consistent with the previous experiment.

The experiment replicates and extends the findings of the previous experiment. The results show that about half of the participants in both experiments tend to look at both the affirmative and the negative image when they hear a counterfactual, or at the negative image; the other half look only at the affirmative image, just as they do for the indicative conditional. In the next experiment we extend the findings to a verbally based visual world paradigm.

EXPERIMENT 3

The aim of the experiment was to test whether the results of the previous experiments are replicated, this time for a verbally based visual world paradigm. The experiment had the same design as the previous experiments, but printed words were shown instead of pictures, as **Figure 1** shows (on the right-hand side). Most of the studies that compare both formats show similar results for them (e.g., McQueen and Viebahn, 2007; Primativo et al., 2016). The printed word version may be more sensitive to phonological manipulations than the traditional picture version (e.g., Huettig and McQueen, 2011, see also Weber et al., 2007). The printed word version is useful for investigating orthographic processing during speech perception but less so for investigating processing of semantic and conceptual visual-form representations (Salverda and Tanenhaus, 2010; Huettig and McQueen, 2011). However, visual information such as pictures has been found to impede relational and conditional reasoning, as well as reasoning about negation (e.g., Knauff and Johnson-Laird, 2002; Orenes and Santamaría, 2014). Hence our aim was to examine whether the same results occur for the printed word version as for the pictures version of the visual world paradigm. We also aimed to rule out

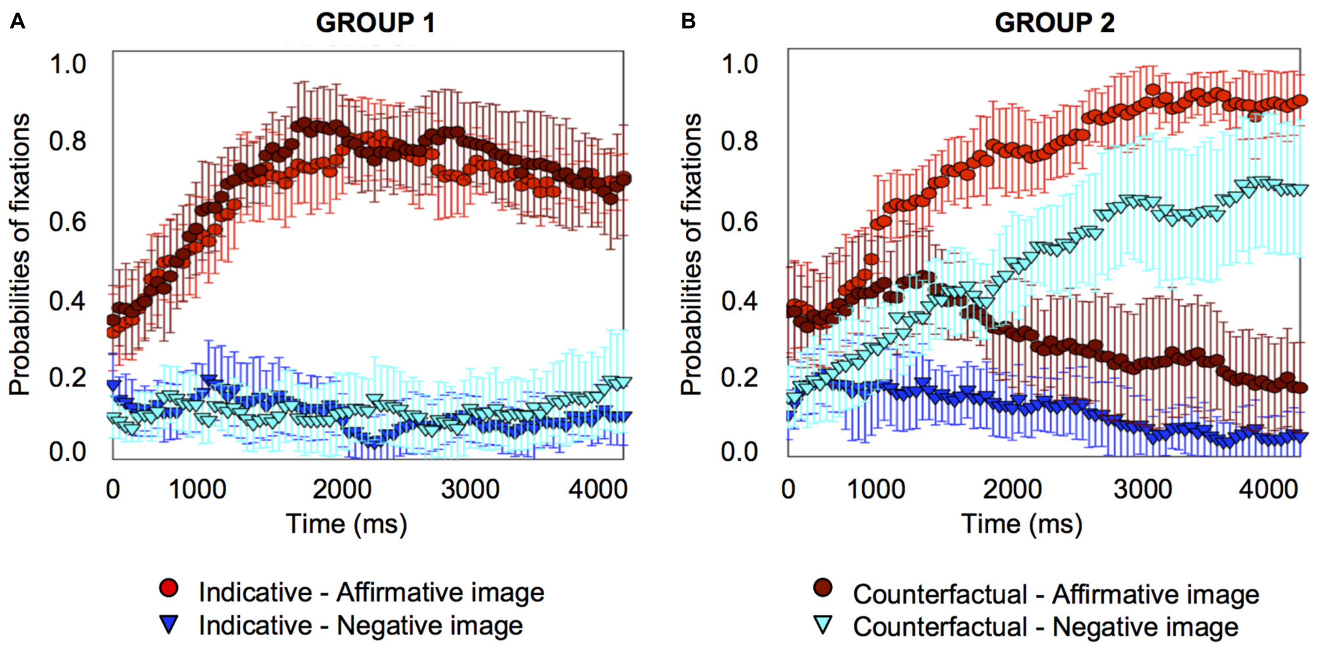


FIGURE 6 | Individual differences probabilities of fixations for indicative and counterfactual sentences for a subset group of 10 participants who looked at the affirmative image for counterfactuals **(A)**, and a subset group of eight participants who looked at the negative image only or at the affirmative and negative image for counterfactuals **(B)** in Experiment 2. Error bars are 95% confidence intervals within participants.

any possibility that the negative visual images used in the previous experiments, e.g., an orange and a pear with a cross through it, was confusing for participants. Participants were explicitly instructed to look at the object or objects on the screen that matched the meaning of the stories that they heard, as they were in Experiment 1.

Methods

Participants

The participants were a new set of 24 native Spanish speakers from the University of La Laguna, Tenerife, Spain, who participated in the experiment in exchange for course credits. There were 22 women and 2 men, and their average age was 20 years, with a range from 18 to 41 years. All of them reported normal vision or wore soft contact lenses or glasses. None of them had taken part in the previous experiments.

Materials, Design and Procedure

The materials, design and procedure were the same as the previous experiments. The only difference was that we presented printed words instead of images on screen, as shown in **Figure 1**. Participants were explicitly instructed to look at the object or objects on the screen that matched the meaning of the stories that they heard, as in Experiment 1.

Results and Discussion

Prior to any data analysis the data of two participants were discarded because one participant fixated on just one point on the screen throughout the experiment and no eye-movements were registered, and the other participant had too many blinks. The procedure for analyzing the eye movement data was the same as that used in the previous experiments.

T-Tests Against Baseline

The results replicated the previous experiments, as **Figure 7** shows. For the indicative conditional, at the onset of the target word, participants were focused on the affirmative printed words and the affirmative distractor equally frequently. From 450 *ms* after the target word onset, the probabilities of fixation on the affirmative printed words started to increase (*pFDR-corr* = 0.027), fixations on the other printed words decreased, including for the negative printed words (from 800 *ms*, *pFDR-corr* = 0.041) (see the **Supplementary Material B** for details about the distractor images). The results replicate the findings of the previous experiments that participants increase their fixation on the affirmative printed words very early on, and fixations on the other three printed words, including the negative printed words, decrease rapidly, as **Figure 7A** shows.

For the counterfactual, at the onset of the target word, participants were focused on the affirmative printed words and the affirmative distractor equally frequently. After the target word onset, fixations on the affirmative printed words remained at about 0.3 to 0.4 and did not change; from 450 *ms* there was an *increase* in fixation on the negative printed words (*pFDR-corr* = 0.039). The results replicate those of the previous experiments that early in the temporal course of processing a counterfactual, within about half a second, participants fixate increasingly

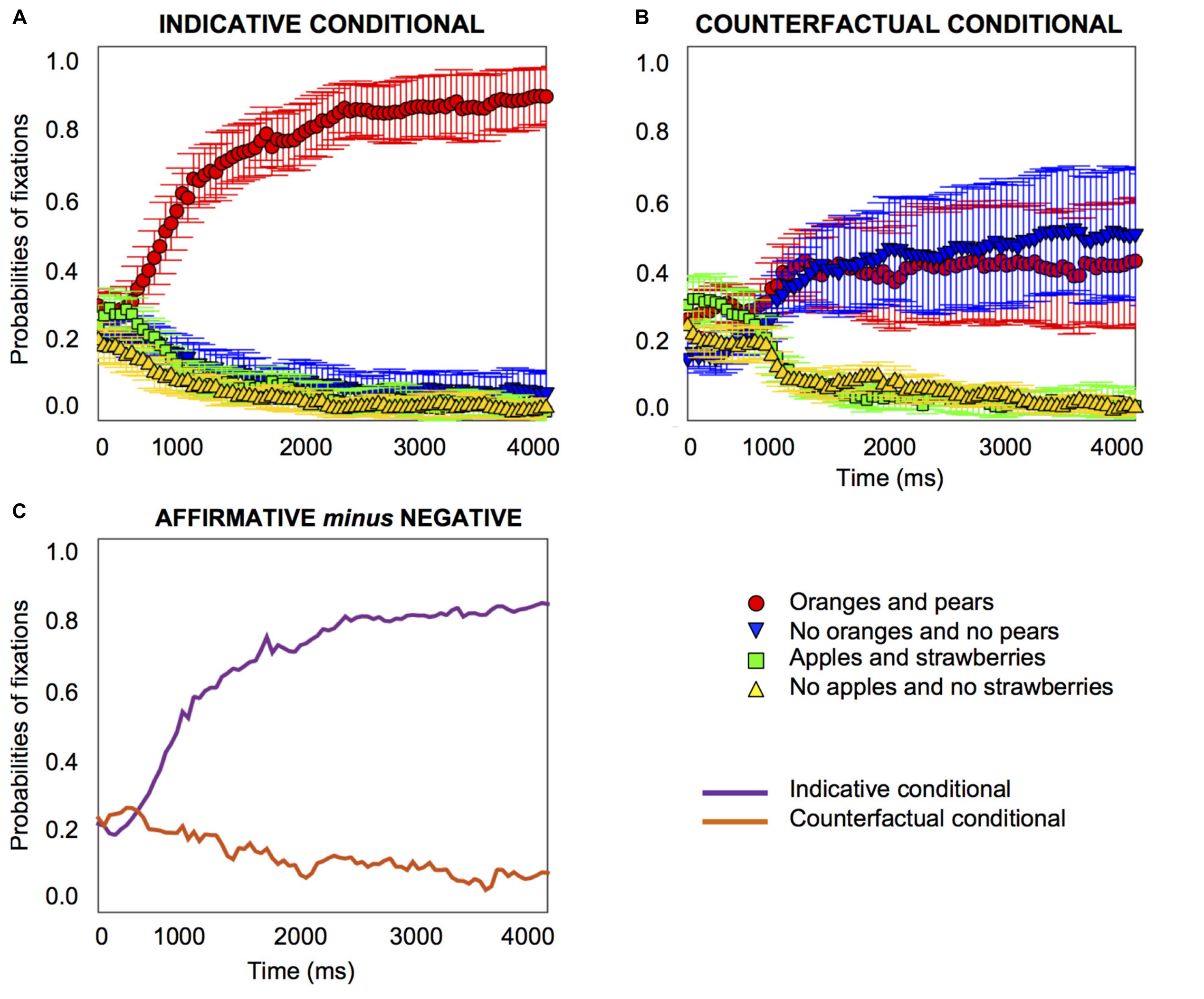


FIGURE 7 | Probabilities of fixations for indicative conditionals, e.g., "if there are oranges, then there are pears" **(A)** and counterfactual conditionals, e.g., "if there had been oranges, then there would have been pears" **(B)** in Experiment 3. The differences in the probabilities of fixations (on the affirmative printed words minus the negative printed words) for indicative and counterfactual conditionals emerge at 550 *ms*, as **(C)** shows. Error bars are 95% confidence intervals within participants.

on the negative printed words, their fixation on the affirmative printed words did not change from the baseline, and fixations on the two distractors decrease rapidly, as **Figure 7B** shows.

Growth-Curve Analysis

The growth curve analysis showed the same results as the previous experiments (see the **Supplementary Material B** for details).

Individual Differences Analysis

The **Supplementary Material C** provides the individual graphs for each of the 22 participants. Almost half of the participants looked at the affirmative printed words only, when they heard the counterfactual (nine participants), just as they did for the indicative conditional; more than half looked at the negative printed words (13 participants, 8 who looked at the negative printed words and 5 who looked at both the negative and the affirmative printed words) as **Figure 8** shows.

For the indicative conditional, both groups showed the same pattern: the probabilities of fixation on the affirmative printed words increased (group 1 from 400 *ms*, *pFDR-corr* = 0.040; group 2 from 600 *ms*, *pFDR-corr* = 0.033) and fixations on the other printed words decreased, including for the negative printed words (group 1 from 1400 *ms*, *pFDR-corr* = 0.032; group 2 from 950 *ms*, *pFDR-corr* = 0.039) (see the **Supplementary Material B** for details about the distractor images). For the counterfactual, the groups showed different patterns. For group 1, the pattern was the same as the indicative conditional, the probabilities of fixation on the affirmative printed words increased (from 500 *ms*, *pFDR-corr* = 0.037), and fixations on the other printed words decreased, including for the negative printed words (from 1850 *ms*, *pFDR-corr* = 0.025). Group 2's pattern was different: probabilities of fixation for the affirmative printed words showed no significant

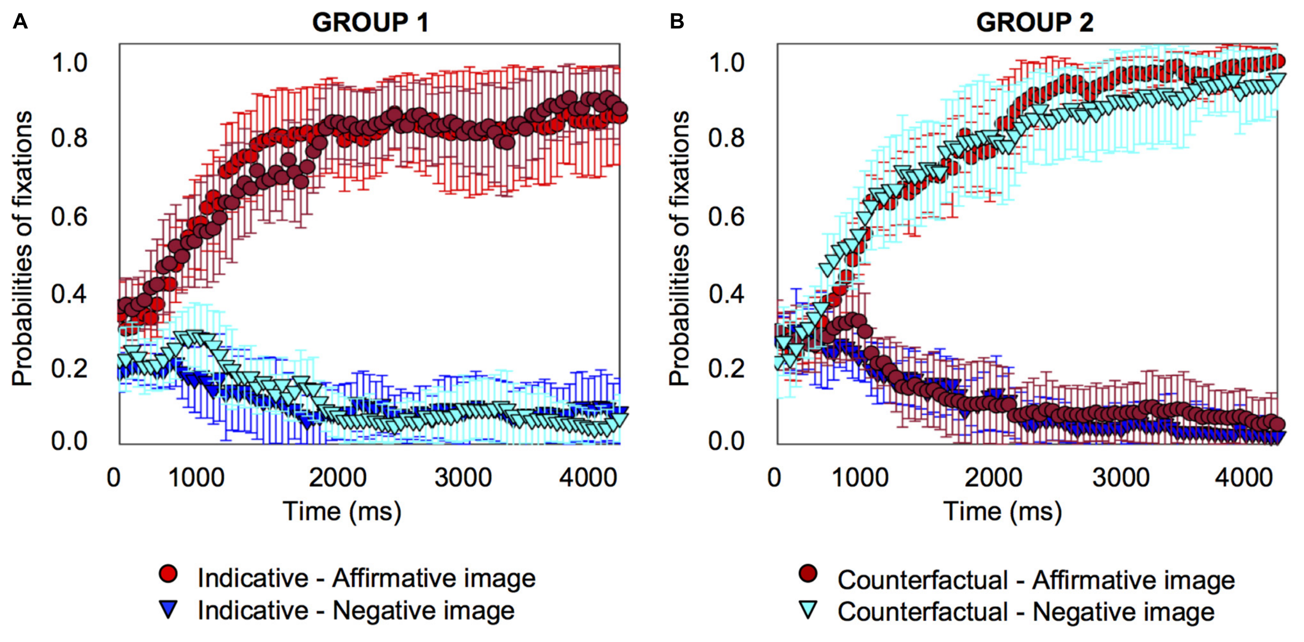


FIGURE 8 | Individual differences probabilities of fixations for indicative and counterfactual sentences for the subset group of nine participants who looked at the affirmative printed words for counterfactuals **(A)**, and the second subset group of 13 participants who looked at the negative printed words only or at the affirmative and negative printed words for counterfactuals **(B)** in Experiment 3. Error bars are 95% confidence intervals within participants.

changes to the baseline, but they *increased* for the negative printed words (from 250 *ms*, *pFDR-corr* = 0.046).

The analysis shows that when people understand the indicative conditional, they look at the affirmative printed words from very early (400–600 *ms*) and decrease their fixations on the negative printed words quite some time later (1400 *ms* in Group 1; 950 *ms* in Group 2). When they understand the counterfactual, one subset of participants do the same thing as for the indicative, they look at the affirmative printed words from very early (500 *ms*) and decrease their fixations on the negative printed words quite some time later (1850 *ms*), but the other subset of participants look at the affirmative printed words early and continue to do so at the same rate as the baseline throughout, but these participants look increasingly at the negative printed words from very early (250 *ms*). The pattern of a subset of participants focusing on the counterfactual negative printed words is particularly clear-cut for the printed word version of the visual world paradigm.

The experiment replicates and extends the findings of the previous experiments when participants are provided with the printed word version of the visual world paradigm. Therefore, the results rule out the possibility that participants were confused in the previous experiments by the representation of the absence of objects, such as "no oranges and no pears," by an image of the objects with a cross through it, or that they experienced other difficulties in identifying the objects. Once again, the overall group data show that when people understand the counterfactual, e.g., "if there had been oranges then there would have been pears," very early in the temporal course of processing, they increase their focus on the negative printed words overall, and continue to maintain their focus on the affirmative printed words at the same rate as at the baseline, and this focus on both sorts of printed words continues throughout the period of measurement. The experiment again shows pronounced individual differences. About half of the participants appear to understand the counterfactual just as they do the indicative conditional, they focus on the affirmative printed words only. The other half of the participants understand the counterfactual differently from the indicative conditional, they focus on the affirmative and the negative printed words.

Individual Differences Analysis Over the Three Experiments

To increase the power of the individual differences analysis, we combined the data from the 63 participants who took part in the three experiments, since they were drawn from the same population. Given that the experiments used the same materials and the results were similar, we carried out an exploratory cluster analysis k-mean to discover similarities in participants' patterns of counterfactual processing. The analysis split participants into two subgroups depending on how they processed counterfactuals. From the combined participant set, 30 participants looked at the affirmative image (or printed words) more frequently than the negative one when they heard the counterfactual (the 11 participants described earlier from experiment 1, 10 from experiment 2, and 9 from experiment 3), and 33 participants looked at the affirmative image (or printed words) less frequently than the negative one (12 participants from experiment 1, 8 from experiment 2, and 13 from experiment 3) as **Figure 9** shows.

For the indicative conditional, both groups showed the same pattern: the probabilities of fixation on the affirmative image (or printed word) increased (group 1 from 350 *ms, pFDR-corr* = 0.008; group 2 from 450 *ms*, *pFDR-corr* = 0.034) and fixations on the other images/printed words decreased, including for the negative image/printed word (group 1 from 1300 *ms*, *pFDR-corr* = 0.034; group 2, from 900 *ms, pFDR-corr* = 0.023) (see the **Supplementary Material B** for details about the distractor images). For the counterfactual, the groups showed different patterns. For group 1, the pattern was the same as the indicative conditional, the probabilities of fixation on the affirmative image (or printed word) increased (from 300 *ms*, *pFDR-corr* = 0.020), and fixations on the other images/printed words decreased, including for the negative image/printed word (from 1200 *ms*, *pFDR-corr* = 0.030). Group 2's pattern was different: probabilities of fixation did not change for the affirmative image/printed word, but they *increased* for the negative image/printed word (from 250 *ms*, *pFDR-corr* = 0.044).

The analysis shows that when people understand the indicative conditional, they look at the affirmative image or printed words from very early (350–450 *ms*) and decrease their fixations on the negative image (or printed word) quite some time later (1300 *ms* in Group 1; 900 *ms* in Group 2). When they understand the counterfactual, one subset of participants do the same thing as for the indicative, they look at the affirmative image or printed words from very early (300 *ms*) and decrease their fixations on the negative image/printed word quite some time later (1200 *ms*), but the other subset of participants look at the affirmative image or printed words early and continue to do so at the same rate as the baseline throughout, and these participants increasingly look at the negative image/printed word and from very early on (250 *ms*). The results show that the previous individual differences analyses hold for the combined larger sample size.

Growth-Curve Analysis

We carried out a growth curve analysis of the combined data from the three experiments (see the **Supplementary Material B** for details). As **Figure 9** shows, for group 1, there were no differences between indicative conditionals and counterfactuals; but for group 2, there were differences for both types of conditional. For the indicative conditional participants in group 2 increase their focus on the affirmative image (or printed word) during the time period measured whereas for the counterfactual conditional they show no significant increase or decrease in their focus on the affirmative image/printed word, and hence they looked at the affirmative image/printed word more for the indicative conditional than for the counterfactual. Moreover, for the counterfactual they increased fixations on the negative image (or printed word) very early, around 250 *ms*, and hence they looked at the negative image/printed word more for the counterfactual than for the indicative conditional. The result confirms the growth curve analyses for each of the three experiments (see the **Supplementary Material B** for details).

GENERAL DISCUSSION

Our objective was to explore the unfolding processing of counterfactual conditionals over time, to test the theory that

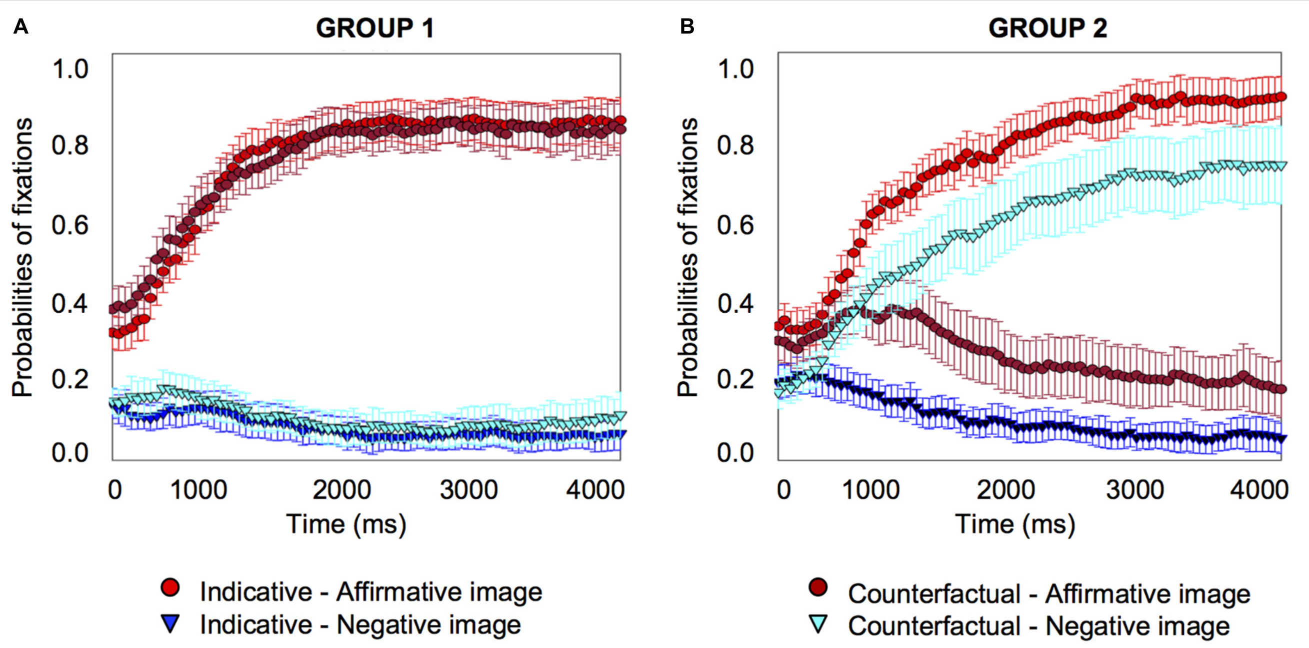


FIGURE 9 | Individual differences probabilities of fixations for indicative and counterfactual sentences for the subset group of 30 participants who looked more frequently at the affirmative image or printed words for counterfactuals **(A)**, and the second subset group of 33 participants who looked more frequently at the negative one for counterfactuals **(B)** in all experiments. Error bars are 95% confidence intervals within participants.

to fully understand a counterfactual conditional, people must imagine the alternative to reality that it conjectures, and they must also recover the known or presumed reality. The three experiments provide information on the temporal course of processing indicative conditionals such as, ‘if there are oranges, then there are pears’ and counterfactual conditionals, such as, “if there had been oranges, then there would have been pears,” by examining the affirmative and negative images that people look at in a visual world display when they hear such conditionals.

We have discovered striking differences in what people look at when they hear counterfactuals and indicative conditionals, as revealed by the analyses of the overall group data in the three experiments. However, we have also discovered notable individual differences in how people understand counterfactuals. For the comprehension of an indicative conditional, e.g., “if there are oranges, then there are pears,” the overall group results reflect the results for the individual participants: each participant exhibited a very rapid focus on the affirmative image or printed word of oranges and pears, within half a second of hearing the target word, e.g., “oranges” (and this focus occurred at about 400 to 450 *ms* in the three experiments). Their focus on the affirmative image or printed word was accompanied by a rapid *decrease* in focus on the negative image or printed word of no oranges and no pears (occurring at about 350 to 800 *ms* in the three experiments).

For the comprehension of a counterfactual conditional, e.g., “if there had been oranges then there would have been pears,” the overall group results reflect the results of one subset of the participants, comprising about half the sample. The overall group data show a very rapid focus, within half a second, on the affirmative image or printed word corresponding to the conjecture, “there were oranges and there were pears” (occurring at about 300–550 *ms* in the three experiments). Strikingly, the focus on the affirmative image or printed word is matched by a rapid *increase* in focus on the negative image or printed word corresponding to the presumed facts, e.g., “there were no oranges and no pears” (occurring at about 450–650 *ms* in the three experiments). Thus, the results from the overall group data indicate that the comprehension of the dual meaning of counterfactuals emerges very rapidly, and participants focus on the affirmative and the negative image within about half a second of hearing the target word, e.g., “oranges” at the end of the antecedent clause. But the individual differences analyses show that these differences are due to one subset of participants. About half of the participants understood the counterfactual differently from the indicative conditional – their focus on the affirmative image or printed word showed no significant increase or decrease from their baseline tendency throughout the time period, but they *increased* their fixations on the negative image or printed word for the counterfactual conditional at a strikingly early time point, from 250 to 700 *ms*. But the other half of the participants in the three experiments appeared instead to understand the counterfactual just as they did the indicative conditional, and they focused only on the affirmative image or printed word (from 350–500 *ms* in the three experiments). They looked only at the affirmative image even for the counterfactual, and they tended to decrease their fixations on the negative image quite late in the temporal process of comprehension (from 1450 to 1850 *ms* in the three experiments).

The results for the overall group data, and for the subset of half of the participants who focused on the negative image or printed word for a counterfactual, corroborate the prediction of a rapid increase in looking at the negative image or printed word as soon as participants detected that the counterfactual communicates an imagined alternative to reality. This finding supports the theory that the recovery of the presumed reality is essential to the full understanding of the meaning of a counterfactual (e.g., Johnson-Laird and Byrne, 1991; Espino and Byrne, 2018). The results also corroborate the prediction of the maintenance of looking at both the negative and the affirmative image or printed word throughout the period of time measured. However, it is notable that the subset of individuals who looked at the negative image or printed word for the counterfactual maintained their focus on the affirmative image or printed word only at a rate similar to their baseline rate, and did not increase their focus on it. The findings from the overall group analysis, and for this subset of half of the participants, support the theory that the essence of the dual meaning of a counterfactual is the comparison of reality to an imagined alternative (e.g., Byrne, 2005; Beck et al., 2006).

The differences between the two types of conditional, when they did emerge, emerged early, as **Figures 2C**, **5C**, **7C** show. They occurred at about 550 to 850 *ms* with explicit instructions to look at the objects that correspond to the meaning of what participants hear (in Experiments 1 and 3), and somewhat later without instructions (at 1700 *ms* in Experiment 2). The instruction (which may activate a controlled or top-down process) accelerates the understanding of counterfactuals in relation to the images. In particular, the differences between the two types of conditionals were due to the increase in attention to the negative image for the counterfactual. The results showed the same pattern for printed words and pictures, which demonstrates the similarities between both methodologies (McQueen and Viebahn, 2007; Primativo et al., 2016). But the tendency to focus on the negative printed word in the third experiment was perhaps even more clear-cut, as **Figures 4**, **6**, **8** illustrate, which is consistent with findings that visual information can impede the comprehension of negation given its symbolic nature (see Orenes and Santamaría, 2014). The findings may also have implications for the question of whether the inference of the falsity of a counterfactual’s antecedent and consequent is a “global” sentential inference accessed only at the end of the sentence (e.g., Sperber and Wilson, 1995), or a “local” sub-sentential inference accessed as soon as some trigger or cue is encountered (e.g., Levinson, 2000; see Reboul, 2004). The early processing of the negative printed word seems to suggest it is not a global inference.

Nonetheless, the data show clearly that almost half of participants did not recover the presumed facts. Why are there such striking differences between individuals in the comprehension of a counterfactual conditional? One explanation could be that they arise from some aspect of the visual world paradigm task. For example, when asked to look at the image or printed word that corresponds to the meaning of the sentence, a participant may interpret that as referring to the way things would have been in the hypothetical situation, that is, the affirmative image or printed word which corresponds to the non-actual situation that the counterfactual sentence invites one to entertain, or to what is implied about actual circumstances, that is, the negative image or printed word that corresponds to the actual circumstances as conveyed by the presupposition of the counterfactual. However, the results of Experiment 1 were replicated in Experiment 2, in which participants were not given explicit instructions to look at the image that corresponded to the meaning of the sentence, and so we can rule out the suggestion that the differences arise from differences in interpretations of the instructions.

Of course, it may also be the case that the visual world paradigm and eye-tracking provides a somewhat insensitive measure of the mental representation of counterfactuals. The objects a person fixates on need not be the only objects they are thinking about, and viewers may even use a broader attentional focus to attend to several objects (e.g., Cave and Bichot, 1999; see also Huettig and Altmann, 2005, see also Huettig and Altmann, 2011; Huettig et al., 2011b). It is also worth noting that participants rarely focused on the distractors, such as the image or printed words corresponding to "apple and strawberry" or "no apple and no strawberry." Strictly speaking, for a counterfactual such as "if there had been oranges there would have been pears," the distractor is also consistent with its presumed facts. For example, the image of an apple and a strawberry can be interpreted as an implicit negation of an orange and a pear (e.g., Espino and Byrne, 2018). Yet, participants focused on the image that contained an explicit negation, the orange and pear with a cross through it, or the printed words "no orange and no pear," rather than on either of the two distractors. It may be that the explicit negation is more salient in the set of four images as corresponding to the opposite of what the counterfactual conjectured, that is, as the presumed facts. Participants may recover the presumed reality from the imagined alternative to reality conjectured in the counterfactual by negating the items mentioned. Of course, it may be more time consuming and require more cognitive steps to make the inference from "no orange and no pear" to "apple and strawberry" (e.g., Espino and Byrne, 2012; Khemlani et al., 2014; Orenes et al., 2014). Moreover, unless the context specifies a binary situation, the inference that "there is no orange and no pear" does not mean necessarily that "there is an apple and a strawberry" since there could be other fruit instead. The lack of attention to the affirmative distractor may arise because during the experimental trials the participants detected that the stories continued after the counterfactual by referring to the items in the affirmative image (e.g., orange and pear) or negative image (e.g., no orange and no pear) in the subsequent conjunction that followed the counterfactual.

An alternative potential explanation for the individual differences is that they arise from difficulties in considering different possibilities for counterfactuals. Such difficulties could arise because of differences in working memory capacity (e.g., Ferguson and Cane, 2015). Participants may focus on only one image as a consequence of limitations of working memory, given that multiple alternatives can overload processing capacity (e.g., Johnson-Laird and Byrne, 2002; Khemlani et al., 2018). Related to this proposal, the differences between individuals may reflect a failure by some participants to process the information deeply. Some participants exhibit a tendency in these sorts of tasks to construct an incomplete and shallow semantic representation (e.g., Ferreira et al., 2002; see also Erickson and Mattson, 1981; Barton and Sanford, 1993). Some participants may represent only the conjecture as a result of a heuristic "match" to what is mentioned in the conditional (e.g., Evans et al., 1999). Of course, the subjunctive mood is neither sufficient nor necessary for the communication of counterfactuality (e.g., Dudman, 1988), and some participants may tend to rely on cues of content more than linguistic mood to trigger a counterfactual interpretation of a conditional. The finding of individual differences in doing so is consistent with inference studies (Thompson and Byrne, 2002). We anticipate that more participants would envisage both the conjecture and the presumed facts for episodic counterfactuals for which the facts are known, compared to the semantic counterfactuals of the current experiments for which the facts must be presumed.

The identification of individual differences in these three experiments has consequences for the interpretation of conflicting observations in previous comprehension studies. The often-conflicting results of previous studies have been interpreted in different ways, either to support the idea that people represent only the conjecture (e.g., Evans and Over, 2004; Evans, 2007), or the idea that they represent both the conjecture and the presumed facts (e.g., Johnson-Laird and Byrne, 2002; Byrne, 2005). Even among theorists who consider that people represent both possibilities, the results have been interpreted to support the idea that the conjecture is more highly activated than the presumed facts (e.g., Nieuwland and Martin, 2012; Kulakova et al., 2013; Ferguson and Cane, 2015), or that the presumed facts are more highly activated than the conjecture (e.g., De Vega and Urrutia, 2012). It seems likely that at least some of the conflicting results reflect individual differences. Nonetheless, the data appear to rule against the theory that people only ever represent a counterfactual's conjecture (e.g., Evans and Over, 2004; Nieuwland and Martin, 2012), since at least half of the participants in each of the experiments represented both the conjecture and the presumed facts. Instead, the data appear to show that a representation of the presumed facts (e.g., the negated conjunction) is a component of the meaning of counterfactuals compared to indicative conditionals, for those participants who reach a counter factual interpretation (e.g., Thompson and Byrne, 2002), and moreover it is very quickly available. The data also suggest, at least for those participants who envisage more than just the conjecture, that the presumed facts may be more highly activated than the conjecture (e.g., De Vega and Urrutia, 2012).

The main contribution of the present study has been to examine the online processing of counterfactual conditionals; hitherto there have been no studies to our knowledge to explore the processing of counterfactuals continuously throughout a 4000 *ms* period of time, measuring eye fixations at every 50 *ms*. Online studies that have explored counterfactuals using event-related potentials (ERP) have focused on one specific period of time (e.g., the component N400; Nieuwland, 2012) and those using eye-tracking have focused on specific intervals (e.g., Stewart et al., 2009). Other eye-tracking or ERP studies of counterfactuals have focused on the effect of the counterfactual conditional on the processing of subsequent words or sentences (e.g., Ferguson et al., 2008; Urrutia et al., 2012a). Similarly, although some studies have highlighted individual differences in the processing of counterfactuals (e.g., Ferguson and Cane, 2015), none has examined different patterns of processing in the focus on affirmative and negative images. The advantage of studying counterfactuals during a continuous 4000 *ms* period and examining fixations at every 50 *ms* is that it has revealed the important discovery that when people hear a counterfactual conditional, about half of them envisage the imagined alternative to reality only, and the other half envisage the imagined alternative to reality and the presumed facts, or the facts alone, and these representational choices occur within just the first few milliseconds after hearing the object word, e.g., "oranges," immediately after the cue of the subjunctive mood, "would have." We chose to time-lock our fixation measurements after the first object (e.g., "oranges") since at that point participants can identify the target images (the ones with oranges and pears or no oranges and no pears) and differentiate them from the distractor images (the ones with apples and strawberries or no apples and no strawberries). By the time participants hear the first object word, however, they have already heard the indicative or subjunctive mood of the antecedent (if there *are*/*had been*), which may provide an additional cue about the likely mood of the consequent (notwithstanding the possibility of mixed antecedent-consequent moods). It would be useful in future studies to explore other time-locks, e.g., after "if." A fruitful avenue for future research may also be to examine individual differences further, to identify their source, and to examine their effects not only on the comprehension of counterfactuals, but also on reasoning with counterfactuals.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Comité de Ética de la Investigación y Bienestar Animal (Universidad de La Laguna)" with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the "Comité de Ética de la Investigación y Bienestar Animal (Universidad de La Laguna)."

AUTHOR CONTRIBUTIONS

IO, JG-M, and RB contributed to the design of the experiments. IO carried out the experiments and analyzed the data. IO and RB wrote the manuscript. All authors contributed to the hypotheses, interpretation of results and revised the manuscript, read and approved the submitted version.

FUNDING

This research was supported by UNED, Grant EDU2014-56423-R, PSI2015-70771-P, Juan de la Cierva Formación and José Castillejo Fellowships from the Spanish Government. Some of the research was carried out while the IO was a visiting researcher at Trinity College Dublin, University of Dublin, Ireland.

ACKNOWLEDGMENTS

We would like to thank Manuel Gutierrez for help with the eye-movement lab at the University of La Laguna Tenerife, and David Beltrán, Gonzalo Duque de Blas, Sabrina Haimovici, Marta Straga, and Shane Timmons for their suggestions and comments. Some of the results were presented at the Interdisciplinary Workshop on Counterfactual Reasoning at the University of Toronto, in Canada in 2016, the Meeting in Memory of Vittorio Girotto in London in 2016, the Workshop on the Philosophy of Cognitive Science in Buenos Aires in 2017, and the Philosophy of Imagination Conference at the Ruhr-University Bochum in 2018.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01172/full#supplementary-material

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

Barton, S. B., and Sanford, A. J. (1993). A case study of anomaly detection: shallow semantic processing and cohesion establishment. *Mem. Cogn.* 2, 477–487. doi: 10.3758/bf03197179

Beck, S. R., Robinson, E. J., Carroll, D. J., and Apperly, I. A. (2006). Children's thinking about counterfactuals and hypotheticals as possibilities. *Child Dev.* 77, 413–423.

Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge: MIT Press.

Byrne, R. M. J. (2016). Counterfactual thought. *Annu. Rev. Psychol.* 67, 135–157. doi: 10.1146/annurev-psych-122414-033249

Byrne, R. M. J., and Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Mem. Cogn.* 27, 726–740. doi: 10.3758/bf03211565

Cave, K., and Bichot, N. P. (1999). Visuospatial attention: beyond a spotlight model. *Psychon. Bull. Rev.* 6, 204–223. doi: 10.3758/bf03212327

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn. Psychol.* 6, 84–107. doi: 10.1016/0010-0285(74)90005-x

De Vega, M., and Urrutia, M. (2012). Discourse updating after reading a counterfactual event. *Psicologica* 33, 157–173.

De Vega, M., Urrutia, M., and Riffo, B. (2007). Cancelling updating in the comprehension of counterfactuals embedded in narratives. *Mem. Cogn.* 35, 1410–1421. doi: 10.3758/bf03193611

Dudman, V. H. (1988). Indicative and subjunctive. *Analysis* 48, 113–122.

Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., and Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: evidence from the visual-world paradigm. *Cognition* 110, 284–292. doi: 10.1016/j.cognition.2008.11.012

Egan, S., and Byrne, R. M. J. (2012). Inferences from counterfactual threats and promise. *Exp. Psychol.* 59, 227–235. doi: 10.1027/1618-3169/a000147

Egan, S., García-Madruga, J. A., and Byrne, R. M. J. (2009). Indicative and counterfactual 'only if' conditionals. *Acta Psychol.* 132, 240–249. doi: 10.1016/j.actpsy.2009.07.005

Erickson, T. A., and Mattson, M. E. (1981). From words to meaning: a semantic illusion. *J. Verbal Learning Verbal Behav.* 20, 540–552. doi: 10.1080/09658211.2016.1152377

Espino, O., and Byrne, R. M. J. (2012). It is not the case that if you understand a conditional you know how to negated it. *J. Cogn. Psychol.* 24, 329–333.

Espino, O., and Byrne, R. M. J. (2018). Thinking about the opposite of what is said: counterfactual conditionals and symbolic or alternate simulations of negation. *Cogn. Sci.* 42, 2459–2501. doi: 10.1111/cogs.12677

Evans, J. S. B. T. (2007). *Essays in Cognitive Psychology. Hypothetical Thinking: Dual Processes in Reasoning and JUDGEMENT*. New York, NY: Psychology Press.

Evans, J. S. B. T., Legrenzi, P., and Girotto, V. (1999). The influence of linguistic form on reasoning: the case of matching bias. *Q. J. Exp. Psychol.* 52, 185–216. doi: 10.1080/027249899391287

Evans, J. S. B. T., and Over, D. E. (2004). *If*. Oxford: Oxford University Press.

Ferguson, H. J. (2012). Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. *Q. J. Exp. Psychol.* 65, 939–961. doi: 10.1080/17470218.2011.637632

Ferguson, H. J., and Cane, J. (2015). Examining the cognitive costs of counterfactual language comprehension: evidence from ERPs. *Brain Res.* 1622, 252–269. doi: 10.1016/j.brainres.2015.05.024

Ferguson, H. J., and Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: an eye-movement investigation. *J. Mem. Lang.* 58, 609–626. doi: 10.1016/j.jml.2007.06.007

Ferguson, H. J., Sanford, A. J., and Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Res.* 1236, 113–125. doi: 10.1016/j.brainres.2008.07.099

Ferguson, H. J., Scheepers, C., and Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Lang. Cogn. Process.* 25, 297–346. doi: 10.1080/01690960903041174

Ferreira, F., Bailey, K. G. D., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Curr. Dir. Psychol. Sci.* 11, 11–15. doi: 10.1111/1467-8721.00158

Fillenbaum, S. (1974). Information amplified: memory for counterfactual conditionals. *J. Exp. Psychol.* 102, 44–49. doi: 10.1037/h0035693

Frosch, C. A., and Byrne, R. M. J. (2012). Causal conditionals and counterfactuals. *Acta Psychol.* 14, 54–66. doi: 10.1016/j.actpsy.2012.07.001

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037

Gómez-Veiga, I., García-Madruga, J. A., and Moreno-Ríos, S. (2010). The interpretation of indicative and subjunctive concessive. *Acta Psychol.* 134, 245–252. doi: 10.1016/j.actpsy.2010.02.005

Huettig, F., and Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition* 96, B23–B32.

Huettig, F., and Altmann, G. T. M. (2011). Looking at anything that is green when hearing frog: how object surface colour and stored object colour knowledge influence language-mediated overt attention. *Q. J. Exp. Psychol.* 64, 122–145. doi: 10.1080/17470218.2010.481474

Huettig, F., and McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Mem. Cogn.* 39, 1068–1084. doi: 10.3758/s13421-011-0086-z

Huettig, F., Rommers, J., and Meyer, A. S. (2011a). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003

Huettig, F., Singh, N., and Mishra, R. K. (2011b). Language-mediated visual orienting behavior in low and high literates. *Front. Psychol.* 2:e285. doi: 10.3389/fpsyg.2011.00285

Johnson-Laird, P. N., and Byrne, R. M. J. (1991). *Essays in Cognitive Psychology. Deduction*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., and Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychol. Rev*. 109, 646–678. doi: 10.1037//0033-295x.109.4.646

Kahneman, D., and Tversky, A. (1982). "The simulation heuristic," in *Judgment Under Uncertainty: Heuristics and Biases*, eds D. Kahneman, P. Slovic, and A. Tversky (New York, NY: Cambridge Univ. Press), 201–208.

Khemlani, S., Byrne, R. M. J., and Johnson-Laird, P. N. (2018). Facts and possibilities: a model based theory of sentential reasoning. *Cogn. Sci.* 42, 1887–1924. doi: 10.1111/cogs.12634

Khemlani, S., Orenes, I., and Johnson-Laird, P. N. (2014). The negations of conjunctions, conditionals, and disjunctions. *Acta Psychol.* 151, 1–7. doi: 10.1016/j.actpsy.2014.05.004

Knauff, M., and Johnson-Laird, P. N. (2002). Imagery can impede inference. *Mem. Cogn.* 30, 363–371. doi: 10.3758/bf03194937

Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. New York, NY: Oxford Univ. Press.

Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., and Perner, J. (2013). Processing counterfactual and hypothetical conditionals: an fMRI investigation. *NeuroImage* 72, 265–271. doi: 10.1016/j.neuroimage.2013.01.060

Kulakova, E., and Nieuwland, M. S. (2016a). Pragmatic skills predict online counterfactual comprehension: evidence from the N400. *Cogn. Affect. Behav. Neurosci.* 16, 814–824. doi: 10.3758/s13415-016-0433-4

Kulakova, E., and Nieuwland, M. S. (2016b). Understanding counterfactuality: a review of experimental evidence for the dual meaning of counterfactual. *Lang. Linguist. Compass* 10, 49–65. doi: 10.1111/lnc3.12175

Levinson, S. (2000). *Presumptive Meaning. The Theory of Generalized Conversational Implicature*. Cambridge, MA: The MIT Press.

Lewis, D. (1973). *Counterfactuals*. London: Blackwell.

Martin, E., Shao, K. C., and Boff, K. R. (1993). Saccadic over-head: information-processing time with and without saccades. *Percep. Psychophys.* 53, 372–380. doi: 10.3758/bf03206780

McQueen, J. M., and Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Q. J. Exp. Psychol.* 60, 661–671. doi: 10.1080/17470210601183890

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: Chapman and Hall. doi: 10.1080/17470210601183890

Moreno-Ríos, S., García-Madruga, J. A., and Byrne, R. M. J. (2008). Semifactual 'even if' reasoning. *Acta Psychol.* 128, 197–209. doi: 10.1080/17470218.2016.1172098

Morey, R. D. (2008). Confidence intervals from normalized data: a correction to cousineau (2005). *Tutorial Q. Methods Psychol.* 4, 61–64. doi: 10.20982/tqmp.04.2.p061

Nieuwland, M. S. (2012). Establishing propositional truth-value in counterfactual and real-world contexts during sentence comprehension: differential sensitivity of the left and right inferior frontal gyri. *NeuroImage* 59, 3433–3440. doi: 10.1016/j.neuroimage.2011.11.018

Nieuwland, M. S. (2013). "If a lion could speak ...": online sensitivity to propositional truth-value of unrealistic counterfactual sentences. *J. Mem. Lang.* 68, 54–67. doi: 10.1016/j.jml.2012.08.003

Nieuwland, M. S., and Martin, A. E. (2012). If the real world were irrelevant, so to speak: the role of propositional truth-value in counterfactual sentence comprehension. *Cognition* 122, 102–109. doi: 10.1016/j.cognition.2011.09.001

O'Brien, F., and Cousineau, D. (2014). Representing error bars in within-subject designs in typical software packages. *Tutorial Q. Methods Psychol.* 10, 56–67. doi: 10.20982/tqmp.10.1.p056

Orenes, I., Beltrán, D., and Santamaría, C. (2014). How negation is understood: evidence from the visual world paradigm. *J. Mem. Lang.* 74, 36–45. doi: 10.1080/17470218.2015.1063675

Orenes, I., Moxey, L., Scheepers, C., and Santamaría, C. (2015). Negation in context evidence from the visual world paradigm. *Q. J. Exp. Psychol.* 69, 1082–1092. doi: 10.1080/17470218.2015.1063675

Orenes, I., and Santamaría, C. (2014). Visual content of words delays negation. *Acta Psychol.* 153, 107–112. doi: 10.1016/j.actpsy.2014.09.013

Primativo, S., Sebastian, J. R., and Crutch, J. (2016). Conceptual Feature ratings predict gaze within written word arrays: evidence from a visual Wor(l)d paradigm. *Cogn. Sci.* 41, 659–685. doi: 10.1111/cogs.12348

Quelhas, A. C., and Byrne, R. M. J. (2003). Reasoning with deontic and counterfactual conditionals. *Think. Reason.* 9, 43–66.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037//0033-2909.124.3.372

Reboul, A. (2004). *Conversational Implicatures: Nonce or Generalized? In Experimental Pragmatics*. London: Palgrave Macmillan, 322–333.

Salverda, A. P., and Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1108–1117. doi: 10.1037/a0019901

Santamaría, C., Espino, O., and Byrne, R. M. J. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *J. Exp. Psychol.: Learn. Mem. Cogn.* 31, 1149–1154. doi: 10.1037/0278-7393.31.5.1149

Sperber, D., and Wilson, D. (1995). *Relevance: Communication and Cognition*, 2nd Edn. Oxford: Basil Blackwell.

Stalnaker, R. C. (1968). "A theory of conditionals," in *Studies in Logical Theory*, ed. N. Rescher (Oxford: Basil Blackwell).

Stewart, A. J., Haigh, M., and Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *Q. J. Exp. Psychol.* 62, 2113–2125. doi: 10.1080/17470210902973106

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863

Thompson, V., and Byrne, R. M. J. (2002). Reasoning counterfactually: making inferences about things that didn't happen. *J. Exp. Psychol. Learn., Mem. Cogn.* 28, 1154–1170. doi: 10.1037//0278-7393.28.6.1154

Urrutia, M., de Vega, M., and Bastiaansen, M. (2012a). Understanding counterfactuals in discourse modulates ERP and oscillatory gamma rhythms in the EEG. *Brain Res.* 1455, 40–55. doi: 10.1016/j.brainres.2012.03.032

Urrutia, M., Gennari, S., and de Vega, M. (2012b). An fMRI study of counterfactual sentences describing physical effort. *Neuropsychologia* 50, 3663–3672. doi: 10.1016/j.neuropsychologia.2012.09.004

Weber, A., Melinger, A., and Lara Tapia, L. (2007). "The mapping of phonetic information to lexical presentations in Spanish: Evidence from eye movements," in *Proceedings of the 16th International Congress of Phonetic Sciences* (ICPhS 2007), eds J. Trouvain and W. J. Barry (Saarbrücken: Saarbruücken University des Saarlandes), 1941–1944.

Williamson, T. (2007). Philosophical knowledge and knowledge of counterfactuals. *Grazer Philos. Stud.* 74, 89–124.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 June 2019
doi: 10.3389/fpsyg.2019.01400

It's Not What You Expected! The Surprising Nature of Cleft Alternatives in French and English

Emilie Destruel[1]*, *David I. Beaver*[2] *and Elizabeth Coppock*[3]

[1] *Department of French and Italian, University of Iowa, Iowa City, IA, United States,* [2] *Department of Linguistics, University of Texas at Austin, Austin, TX, United States,* [3] *Department of Linguistics, Boston University, Boston, MA, United States*

While much prior literature on the meaning of clefts—such as the English form "it is X who Z-ed"—concentrates on the nature and status of the exhaustivity inference ("nobody/nothing other than X Z"), we report on experiments examining the role of the doxastic status of alternatives on the naturalness of *c'est*-clefts in French and *it*-clefts in English. Specifically, we study the hypothesis that clefts indicate a conflict with a doxastic commitment held by some discourse participant. Results from naturalness tasks suggest that clefts are improved by a property we term "contrariness" (along the lines of Zimmermann, 2008). This property has a gradient effect on felicity judgments: the more strongly interlocutors appear committed to an apparently false notion, the better it is to repudiate them with a cleft.

Keywords: English, French, clefts, contrast, interlocutors' expectations, existential inference

OPEN ACCESS

Edited by:
Sophie Repp,
University of Cologne, Germany

Reviewed by:
Mila Vulchanova,
Norwegian University of Science and Technology, Norway
Clemens Mayr,
University of Göttingen, Germany
Tue Trinh,
Leibniz-Zentrums Allgemeine Sprachwissenschaft (ZAS), Germany

****Correspondence:***
Emilie Destruel
e-destruel-johnson@uiowa.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *30 October 2018*
Accepted: *29 May 2019*
Published: *18 June 2019*

Citation:
Destruel E, Beaver DI and Coppock E (2019) It's Not What You Expected! The Surprising Nature of Cleft Alternatives in French and English. Front. Psychol. 10:1400. doi: 10.3389/fpsyg.2019.01400

1. INTRODUCTION

In many languages, a sentence expressing a single proposition can be cleft in twain, dividing the message over two clauses. Two examples are the English *it*-cleft (1-a) and its French counterpart the *c'est*-cleft (1-b).

(1) a. It's [David]$_F$ who drank vodka.
b. C'est [David]$_F$ qui a bu de la vodka.

It is generally accepted that one purpose that clefting serves is to mark focus. Focus-marking entails that there are alternatives relevant for interpretation, and that those alternatives correspond to the focus-marked constituent (see e.g., Rooth, 1992; Krifka, 2008). In (1), the focus-marked constituent is the so-called *pivot* of the cleft, corresponding to the subject of the embedded clause in this case, and the alternatives correspond to other people who could have drank vodka, e.g., Paul, Jill, etc.

This paper investigates a relatively under-explored aspect of the focal alternatives determined by a cleft, namely their doxastic status for the interlocutor. In particular, we investigate the possibility that clefts signal a commitment on the part of the interlocutor to a proposition that conflicts with the one the cleft expresses, and that clefts serve to express opposition to that commitment. Using acceptability rating tasks, we provide experimental evidence that, *ceteris paribus*, both *it*-clefts and *c'est*-clefts improve in acceptability in proportion to the degree to which they indicate that an utterance runs contrary to a doxastic commitment on the part of the interlocutor (or another discourse participant).

Clefts have generally been analyzed as conveying three types of information (Halvorsen, 1978; Horn, 1981; Lambrecht, 1994), which we will refer to as the *Halvorsen components*. The first is the at-issue content, often referred to as the PREJACENT, which for a sentence of the form "it is X who Z-ed" is the proposition expressed by the canonical form "X Z-ed" (2-a). Second, clefts convey an EXISTENTIAL inference, such that there exists an X who Z-ed (2-b). Unlike the prejacent, this aspect of clefts is typically taken to be a presupposition. Third, they convey an EXHAUSTIVE inference such that X is the sole (or maximal) entity for which Z holds, e.g., (2-c)—the exact nature of which is still a matter of debate (see among others, Halvorsen, 1978; Atlas and Levinson, 1981; Wedgewood, 2007; Velleman et al., 2012; Büring and Kriz, 2013; Destruel et al., 2015). We briefly note for concreteness that we simply assume clefts have identical semantics as the prejacent, as argued for instance in Horn (1981)[1].

(2) a. Prejacent: David drank vodka.
b. Existential: Someone drank vodka.
c. Exhaustivity: No one other than David drank vodka.

The doxastic status of these or other inferences for the hearer has typically not been discussed *p*er se, although Prince (1978) is one exception. On the basis of corpus evidence, she concluded that although *it*-clefts mark the existential as a "known fact," yet "the information represented in *it*-cleft that-clauses does NOT have to be assumed to be in the hearer's mind." Thus while the existential inference is presupposed, it can be what she terms an "informative presupposition." She goes even further, claiming that "*it*-clefts make no assumptions about the hearer." This latter claim is challenged by the data we present here.

Clefts have also been claimed to express *contrast* (Jespersen, 1927; Harries-Delisle, 1978; Sarnicola, 1988; Umbach, 2004; Patten, 2012). For English, this observation dates back to the work in which the term "cleft" was first coined; in perhaps the first general treatment of clefts, Jespersen (1927) claims "A cleaving of a sentence by means of *it is* (often followed by a relative pronoun or connective) serves to single out one particular element of the sentence and very often, by directing attention to it and bringing it, as it were, into focus, to mark a contrast (Jespersen, 1927, 147f.). For French, a similar observation is found in the seminal work of Lambrecht (1994), who argues that the *c'est*-cleft is the most natural way to signal *contrastive focus*, a type of focus that is sometimes distinguished from *information focus* (see e.g., Zimmermann and Onea, 2011). The former signals contrast, while the latter highlights new information.

How is contrast defined? On a broad view, adopted in the work of Vallduvi and Vilkuna (1998) (see also Selkirk, 2008; Lopez, 2009; Katz and Selkirk, 2011), a *kontrastive* expression *a* generates a membership set $M = \{..., a, ...\}$ which "becomes available to semantic computation as some sort of quantificational domain" (Vallduvi and Vilkuna, 1998). Contrast (formalized as *kontrast*) amounts to nothing more than membership in a salient set on this understanding. We note that this definition of contrast corresponds exactly to the definition of focus in Rooth (1985)'s Alternative Semantics in that the contrastive element generates a set of alternatives for the focused constituent.

Several more narrow conceptions of contrast exist as well. Rooth (1992) defines contrast as a subcase of a more general notion of focus; for him, a phrase α should be taken as contrasting with a phrase β if the ordinary semantic value of β is an element of the focus semantic value of α. É. Kiss (1998) writes that focus (for which she uses the term "identificational" focus) has the feature [+contrastive] "if it operates on a closed set of entities whose members are known to the participants of the discourse [...]. In this case, the identification of a subset of the given set also identifies the contrasting complementary subset" (p. 267). This definition requires more than the broad one in that the set of alternatives to the focal element must also be *restricted* in size, and clearly *identifiable* by the discourse participants. Contrast has also been characterized with a requirement to exclude alternatives (Molnár, 2002; Kenesei, 2006); in other words, contrast entails *exhaustivity* on this view. Both Rooth's and Kiss's conceptions of contrast entail a requirement for a salient antecedent in the discourse, a requirement that goes beyond the three Halvorsen components ordinarily attributed to clefts.

There is some evidence that contrast in one of these narrower conceptions is indeed encoded by clefts, as they do appear to require a salient antecedent. For instance, while *it*-clefts often sound odd as direct answers to overt questions as in (3)—i.e., when there is no antecedent in the discourse—they are often much more natural as corrections, as in (4). In this case, the previous utterance being corrected provides exactly the kind of antecedent that Rooth mentions for a contrastive focus.

(3) A: Who cooked the beans?
B: #It was John who cooked the beans[2].

(4) A: I wonder why Alex cooked so much beans.
B: Actually, it was John who cooked the beans.

Quantitative evidence that this contrast is robust comes from Destruel and Velleman (2014), who find that the context in (3) leads to the lowest naturalness ratings for clefts. If clefts encode contrast in Rooth's or Kiss's sense, then these differences can be explained[3].

[1]We further note that our assumptions about the truth-conditional semantics, and indeed about the contrariness implications, are compatible with various explanations given in the literature, which derive the existential and exhaustive implications either by adding presuppositions or through a separate pragmatic process, see for instance (Horn, 1981; Destruel, 2013). Specifically, the conclusions we argue for with regard to contrariness are compatible with a QUD analysis as proposed by e.g., Velleman et al. (2012), but the contrariness results show that the constraints on the QUD would need to be strengthened to account for the requirement that there is a salient contrary view on the correct answer to that expressed by the prejacent.

[2]Throughout the paper, we will indicate ungrammaticality with an asterisk (*) and infelicity with a hash (#).

[3]For É. Kiss (1998), contrast is directly encoded in grammar and realized in a specific sentence position, via clefting in English, and in the left periphery in Hungarian. Put slightly differently, contrast is conventionally encoded in the cleft structure itself. Other scholars, however, argue against the existence of such a strong link, including Horn (1981) and Declerck (1984). On the basis of experimental data in Georgian, Skopeteas and Fanselow (2010a) argue that

But even if the three Halvorsen components introduced in (2) are supplemented with a requirement for contrast in Rooth's or Kiss's sense (i.e., a requirement for the right sort of antecedent), the resulting theoretical picture still fails to capture certain facts about cleft behavior. In English, in contexts where an appropriate discourse-familiar alternative is indeed available, speakers may nevertheless choose *not* to use a cleft—its use sounding stilted and odd. Although experimental work on contrast in clefts is scarce, in a study conducted by Destruel and Velleman (2014), English speakers displayed a statistically robust preference for the canonical version in (5). They also rated the sentence in (5-b) as less natural than (6-b), despite the fact that (5-b) *does* have an antecedent available (viz. Canada), and (6-b) does not[4].

(5) A: Darren sounded really excited about his vacation. I think he might be going to Canada.
  a. B: Actually, he's going to Mexico.
  b. B: ? Actually, it's Mexico that he's going to.

(6) A: We were planning Amy's surprise party for weeks. I can't believe she found out about it. Who told her about it?
  a. B: Ken told her about it.
  b. B: It was Ken who told her about it.

Finally, it is worth noting that there is evidence that in certain languages, clefts or other intuitive contrastive focus constructions do not always lead to the exclusion of alternatives; the strength of the exhaustive inference can in fact be modulated by the context. This has been argued, for instance, for clefts in St'át'imcets (Salish; Thoma, 2009) and French (Destruel and DeVeaugh-Geiss, 2018), for focus movement structures in K'ichee' which are arguably clefts (Mayan; Yasavul, 2013), and for non-cleft focus movement structures in Tangale (Chadic; Zimmermann, 2011) which, Zimmerman argues, still show signs of being contrastive in an important sense. Thus, if we want to retain the idea that clefts (and other focus movement constructions) are inherently contrastive, then these data suggest that defining contrast in terms of exclusion of alternatives may also miss the mark.

Given this backdrop, the question arises as to whether another factor might be relevant in better predicting the clefts' use in contrastive focus contexts, and more broadly, in characterizing the notion of contrast. We think that an interesting approach is found in Zimmermann (2008) and Zimmermann (2011), who proposes a definition calling on the notion of speaker-hearer *expectation*. This definition can therefore be thought of as *doxastic*. A focus constituent α is contrastive whenever the speaker assumes that "the hearer will not consider the content of α or the speech act containing α likely to be(come) common ground" (Zimmermann, 2008, 9). This suggestion is consonant with an earlier claim of Delin (1991), based on an extensive corpus study, that one of several different functions of *it*-clefts is "to correct some previous claim by challenging it." Thus, our first research question is the following:

(7) **Research Question 1**
What factor(s) other than the presence of a discourse-familiar alternative licenses clefts, and, specifically, does the attitude expressed toward salient alternatives affect the felicity of clefts?

Our research on this question builds on previous work by augmenting traditional analyses of contrast with what we term *contrariness*. In the spirit of Zimmermann (2008) and Zimmermann (2011), we take contrariness to relate to the degree of commitment that an addressee is established to have to a contrary focal alternative. More specifically, in our view, contrariness has the following three properties: (a) **contrast** (contrariness of one utterance in the discourse requires another utterance such that the first is an element of the alternative set of the second), (b) **contradiction** (taken together the two utterances are inconsistent, i.e., they entail falsity), (c) **strength** or degree of contrariness (which monotonically increases with the degree of commitment of the speaker to inconsistent propositions expressed by these utterances). We then distinguish between three imaginable hypotheses:

(i) The meaning components identified by Halvorsen (1978) (**the Halvorsen components**) are sufficient to capture the significance of a cleft construction. The contribution of alternatives to the meaning of a cleft lies solely in the exhaustivity component of the meaning.
(ii) In addition to the Halvorsen components, clefts signal a **non-doxastic** type of contrast, of the type characterized by É. Kiss (1998) or Rooth (1992), incorporating a requirement for an appropriate antecedent.
(iii) In addition to the Halvorsen components, clefts signal a **doxastic** type of contrast (i.e., *contrariness*). The nature of the clefted alternatives involves a contrast between interlocutors' expectations[5].

The experiments reported in this paper set out to test Hypothesis (iii)—we hypothesize that in addition to the core components in (2), clefts incorporate a requirement that the ordinary meaning is contrary to a previously salient focal alternative. Put slightly differently, we expect clefts to be optimal candidates in contexts where they do more than just introduce a linguistic contrast, but rather are used as a response to an (explicit) contrary claim. We expect this effect to be gradient on felicity judgments: the more strongly interlocutors appear committed to an apparently false notion, the better it is to repudiate them with a cleft. Crucially, this doxastic definition allows for *degrees* of contrast, corresponding to stronger or weaker conflict with expectations, and we argue that these degrees correlate with clefts' naturalness. On this basis, the slight infelicity of (5-b) might be explained as follows: Although there is some contrast between B's claim and

contrast-related movements are optional, providing evidence that foci realized in the pre-verbal or post-verbal position can receive the same kinds of interpretations (e.g., contrastive or exhaustive). This suggests that contrastivity is not directly encoded in the grammar, at least in Georgian.

[4]Using a 5-point Likert scale, Destruel and Velleman (2014) found that the mean ratings was 1.7 for (5-b), and 2.3 for (6-b).

[5]We discuss how our notion of *contrariness* relates to (accounts of) correctivity in section 3.

what A has stated previously, A's hedging ("I think he might…") indicates only a mild commitment to a contrary proposition, and this mild commitment to a contrary proposition does not suffice to make a cleft fully felicitous for B. Compare this with the much more strident rebuttal of what the hearer suggests is some people's view found in this naturally occurring example cited by Hedberg (1990)[6]:

(8) JM: Some people think that Reagan's administration is at its LOWEST ebb, its NADIR. Do you agree, Eleanor?
EC: Absolutely not. The Reagan-Baker Administration is in FINE shape. It's the BUCHANAN administration that's having PROBLEMS.

A second issue central to our current research concerns the grammatical reflex of contrast across languages. Indeed, while the bulk of the past theoretical literature on focus and clefts has been developed around (introspective judgments for) English, cross-linguistic counterparts to the *it*-cleft are also noted to express contrast, such as the French *c'est*-cleft, as mentioned earlier. But, as Repp (2016) notes, languages might differ with respect to the grammatical sensitivity they have to particular aspects of the (set of) alternatives. The author says that "for instance, the view that alternativeness equals contrastiveness might make the right prediction for the application of particular strategies in language *x* whereas in language *y* similar marking strategies might require the presence of a clearly identifiable alternative set." This seems particularly relevant when comparing clefts in languages like French and English since, while both it- and *c'est*-clefts can express contrast, there are subtle and crucial differences in their distribution. First, the French cleft is used more commonly than its English counterpart (Carter-Thomas, 2009), in particular in comparison to canonical sentence forms (SVO). The reason appears to be primarily prosodic: whereas English can shift prosodic prominence to match the location of the focus constituent, French is more rigid, and prosodic stress is required to appear at the right edge of an intonation phrase. The *c'est*-cleft, despite adding syntactic complexity, circumvents this prosodic restriction by creating an extra intonational boundary that can align with the focus constituent (Hamlaoui, 2008). Consequently, the *c'est*-cleft constitutes the default strategy to signal focus (also known as *information* focus), especially on grammatical subjects (Lambrecht, 1994; Destruel, 2013; Féry, 2013). Second, the French *c'est*-cleft can be used in focus contexts where the English cleft is prohibited; for instance to signal focus on the entire sentence rather than on a single element (i.e., *broad* focus). Given that clefts have a broader distribution in French than English, our paper seeks to address a second research question:

(9) **Research Question 2**
Does dependency of the status of alternatives differ between these two languages?

[6]Note that we are simplifying in the current paper by only considering cases where the speaker disagrees with the addressee, but a more general definition of contrast would allow for disagreement with third parties.

Our research on the second question builds on prior work by directly comparing the role of contrariness in two languages that have different use-conditions for the cleft construction. Given the subtle differences in clefts' use in French and English, we expect that the two languages may differ as to how contrastive a discourse must be before the cleft is considered most natural.

To the best of our knowledge, there have been very few attempts to investigate the contrastive aspect of clefts experimentally (but see Destruel and Velleman, 2014), and especially across languages that differ in their use of clefting as a strategy to mark focus. Moreover, in attempts that do exist, contrast is not often operationalized in a gradient way, i.e., studies typically compare highly contrastive contexts to non-contrastive ones, leaving aside the potential different degrees that contrast can have. Given these observations and the background information presented thus far, this paper aims to bridge the theoretical and the empirical literature on contrast in clefts. The remainder of the paper is structured as follows: We present the studies in section 2, discuss their results in light of current views of contrast and correctivity, and clefts' meaning in section 3. We end with concluding remarks as well as avenues for future work in section 4.

2. THE STUDIES

Recall that the paper examines two research questions, repeated in (10) and (11) for convenience.

(10) **Research Question 1**
What factor(s) other than the presence of a discourse-familiar alternative licenses clefts, and, specifically, does the attitude expressed toward salient alternatives affect the felicity of clefts?

(11) **Research Question 2**
Given that clefts have a broader distribution in French than in English, does dependency of the status of alternatives differ between these two languages?

Our investigation includes three tasks conducted in English and French. Two pre-tests were designed to provide baseline ratings for the existential inference in target sentences and for the strength of commitment of Speaker A in the context, respectively; the main task consisted of naturalness ratings for clefts and canonical sentences in six contexts that instantiated different degrees of contrariness. The experimental stimuli for these three tasks were always presented in written form and were based off of the same source sentences, which were translated by a French native speaker for the French version of the experiment. What differed across tasks regarding the materials was which part of the stimuli participants got to see and judge. Given this, we present the common elements of the three tasks in section 2.1. We present the details for each task—i.e., design, procedure and results—in sections 2.2–2.4.

2.1. Methods

2.1.1. Materials

The experimental stimuli consisted of short dialogues between two speakers. All dialogues included a **background** (Speaker A) as in (12), and a **comment** (Speaker B) presented either in a canonical SVO or in a cleft form, as in (13). Note that the sample stimuli in (12)–(13) illustrate the condition in which the focus is on the grammatical subject. See (14) for an example of the object condition, and **Appendix A** for a larger sample of stimuli. The background always contained three sentences. The first two established the story and the last one contained the information on which B's comment was based. The last sentence in Speaker A's part was crucial in our experiment; this is the sentence we modulated to create six contexts with varying degrees of contrariness, illustrated in (12-a)-(12-f). These six contexts varied according to four factors: Grammatical Function, Contradiction, Commitment and At-issueness. We detail them individually hereafter. For each of the six contexts, we created 12 lexicalizations, so 72 experimental dialogues per grammatical function or 144 in total, and this for each language.

(12) Speaker A: We were planning Amy's surprise party for weeks. I can't believe she found out about it. [...]

- a. Non-contradictory, At-issue (NO CONTR.)
  ... I guess someone from the staff told her.
- b. Weak, At-issue (WEAK)
  ... I guess Alice must have told her.
- c. Weak, Non-At-issue (WEAK NAI)
  ... And Alice—who I think, probably went and told her about it—just laughed and said it was no big deal!
- d. Strong, At-issue (STRONG)
  ... Alice told her about it, you know.
- e. Strong, Non-At-issue (STRONG NAI)
  ... And Alice—who went and told her about it—just laughed and said it was no big deal!
- f. Strong Presuppositional, Non-At-Issue (STRONG PRE.)
  ... I'm annoyed that Alice told her about it!

(13) Speaker B: Yeah/ Actually, [...]

- a. ... Ken told her about it. *(canonical form)*
- b. ... it's Ken who told her about it. *(cleft form)*

The first factor varied was GRAMMATICAL FUNCTION of the focused element, that is whether the element that B commented on was the grammatical subject or the object. Example (14) illustrates the object condition for context NO CONTR.

(14) Object condition, NO CONTR.:

- a. Speaker A: Look at John this evening! He's all dressed up. [...] I guess he's going out with someone from the marketing team.
- b. Speaker B: Yeah, he's going out with Karen/ Yeah, it's Karen he's going out with.

The second factor, CONTRADICTION, refers to whether or not the information in Speaker B's comment contradicted the information stated in the last sentence uttered by Speaker A. We can think of this variable as binary: The first context we designed (NO CONTR.) has a contradiction value of 0 (i.e., it is non-contradictory) because there is no other identifiable salient individual in A's part. The other five contexts have a contradiction value equal to 1; they are contradictory in the sense that there is one alternative explicitly given in the discourse, thus being clearly identified. In the non-contradictory context, B's comment was always introduced by "Yeah/*Ouai*,...," while in all others, B's comment was introduced by "Actually/*En fait*,...".

The third factor we manipulated was AT-ISSUENESS, which refers to whether or not the relevant proposition in A's speech commented on by B was at-issue. The motivation behind including AT-ISSUENESS as a factor comes from Destruel and Velleman (2014), who also argue for the relevance of *contrast in expectation* in the interpretation of clefts, and propose that two types of expectations may be at play; not just expectations about the state of the world but also expectations about the shape and direction of discourse. The latter type is directly relevant here since it may involve beliefs about the direction in which the discourse is going, expressed, among other ways, by marking content as at-issue or not-at-issue. We assume that interlocutors taking part in a discourse will generally address the propositions that are currently at-issue.

Finally, we varied COMMITMENT, which corresponds to the strength with which Speaker A is committed to their statement. Expanding on prior studies on the (grammatical) reflexes of contrast, we take this factor to be gradient; it can vary in strength depending on how the speaker chooses to express their beliefs. We designed contexts that varied in ways that we assumed would affect the level of commitment[7]. We used a variety of attitude verbs and adverbs to encode these various degrees. For instance, in the weak and strong conditions (contexts WEAK to STRONG NAI), the speaker respectively expresses a low or a high degree of commitment toward the asserted prejacent proposition. In context STRONG PRE., on the contrary, the prejacent is presupposed; the speaker expresses a personal, subjective opinion about the truth of another asserted proposition in the sentence (i.e., "I'm annoyed that Alice told her about it!," in (12-f)). Since at-issueness reflects differences in whether a speaker has decided to foreground commitment to a proposition, we anticipated that at-issueness might affect the perceived level of commitment. Further, different types of non-at-issue material (conventional implicatures vs. presuppositions) might also be expected to affect perceived commitment in different ways, e.g., because presupposed non-at-issue material is often taken to reflect a shared commitment, whereas other types of non-at-issue material, such as conventional implicatures from parenthetical, are not. Therefore we included both stimuli in which the target proposition was presupposed, and material in which it was conventionally implicated (in the sense of Potts, 2005).

[7] A reviewer asks why subjects should attribute a stronger commitment to A toward the proposition in context WEAK than in context NO CONTR.. We believe they should because in context 2, speaker A attributes a particular value, that is identifies a specific referent, that is then contradicted by B.

A pre-test (task 2), which we detail in section 2.3 was conducted prior to the main task in order to assess whether our contexts were indeed different with respect to the strength of COMMITMENT encoded, as we conceived them to be. In our view, more strongly expressed commitment lead to stronger conflict between interlocutors, and thus we hypothesized that clefts are more natural in cases when the level of conflict between interlocutors is maximal, or in other words, when clefts are used as responses to an (explicit) contrary claim.

We now turn to discussing each task individually, the two pre-tests first (sections 2.2 and 2.3) and then the main task (section 2.4).

2.2. Task 1: Strength of Existential Inference

2.2.1. Participants

We note that all participants in Task 1 were different from the participants who completed Task 2 and the main task.

For English: We recruited a total of 65 participants (all undergraduates at a midwestern university, ages: 19–23; median: 20) from a first-year language class. Subjects were given extra-credit for their participation and were all naive as to the goal of the experiment.

For French: We recruited 48 monolingual native speakers of French. All were given monetary compensation for their participation and were naive as to the goal of the experiment. Participants were from the regions of Pau, Toulouse and Albi in Southwestern France. Overall, 61% were undergraduate students, 34% graduate students, and 5% staff working at the university.

2.2.2. Design & Procedure

The goal of this first test was to measure the strength of the existential inference in Speaker A's part, i.e., **how likely is it that A believes someone "Z-ed"?** This is necessary to ensure that any effect of contrariness we find is not an artifact of variation among items with respect to the strength of the existential inference that they give rise to. The test was delivered via the web-based survey site Qualtrics. Participants sat in front of a computer screen and read a total of 24 backgrounds (A's part), pseudo-randomized among 24 fillers (recall that participants only saw and rated Speaker A's part of the dialogue in this task.) On each trial, after reading A's part, participants were asked to judge, on a scale from 1 to 7, how likely is it that A thinks that someone Z-ed. So for instance, given NO CONTR. context in (12-a) above, participants were asked how likely is it that "A believes someone told Amy about her surprise party" (1 corresponding to extremely unlikely and 7 to extremely likely). The procedure for English and French was exactly similar; French speakers provided judgments based on the question "*Quelle est la probabilité que A pense que quelqu'un a Z?*"

2.2.3. Results

Mean probability ratings for the strength of the existential inference in A's part are presented in **Table 1**, for English and French.

Visual inspection of these averages suggests that participants deem the likelihood of speaker thinking that someone Z-ed lower for the context that lacks a contrast between A's sentence and B's response (i.e., context NO CONTR., μ = 4.5/4.65), vs. other contexts (where μ is consistently above 6.3)—and this quite similarly in both languages.

TABLE 1 | Mean probability judgments for pre-test 1 (Strength of existential inference).

	Mean ratings (subjects)		Mean ratings (objects)		Overall ratings	
	English	**French**	**English**	**French**	**English**	**French**
No contr.	4.6	4.5	4.4	4.8	4.5	4.65
Weak	6.5	6.3	6.3	6.4	6.4	6.35
Weak nai	6.5	6.4	6.5	6.4	6.5	6.4
Strong	6.6	6.4	6.7	6.6	6.7	6.5
Strong nai	6.4	6.8	6.4	6.6	6.4	6.7
Strong pre.	6.7	6.7	6.7	6.8	6.7	6.75

To determine whether participants' existential ratings varied depending on the fixed-effect predictor CONTRAST (sum-coded prior to analysis as -1/1 for context (NO CONTR. vs. others, respectively), we fit a linear mixed effect model to the data for each language. The two models included the maximal random effects structure justified by the data: random by-item intercepts, random by-participant intercepts and random slopes for CONTRAST by item and participant. *P*-values were obtained by likelihood ratio test of the full model with the effect in question against the model without the effect in question. Results reveal a significant effect of CONTRAST both in English (β = 2.043, *SE* = 0.091, t = 22.24, p < .001), and French (β = 1.62, *SE* = 0.24, t = 6.72, p < .001) suggesting that, as expected, there was a difference in ratings between the non-contrastive context (#1, in (12-a)) vs. the others where a conjecture was present [contexts in (12-b)–(12-f)].

Crucially though, when looking only at the contradictory contexts, we see that the ratings do not significantly differ from each other with respect to A's commitment to existence. This is an important finding since it indicates homogeneity across these contexts. If we also find that these contexts differ in the strength of A's commitment to a statement that B will contradict (as they were designed to do and is tested in task 2), then we will be able to test our prediction that clefts' naturalness is best predicted by a doxastic contrast (i.e., Hypothesis iii.).

2.3. Task 2: Strength of Commitment

2.3.1. Participants

We note that all participants in Task 2 were different from the participants who completed Task 1 and the main task.

For English: We recruited a total of 65 participants (all undergraduates at a midwestern university, ages: 18–21; median: 20) from a first-year language class. Subjects were given extra-credit for their participation and were all naive as to the goal of the experiment.

For French: We recruited 48 monolingual native speakers of French. All were given monetary compensation for their participation and were naive as to the goal of the experiment.

Participants were from the regions of Pau, Toulouse and Albi in Southwestern France. Overall, 83% were undergraduate students, 15% graduate students, and 2% staff working at the university.

2.3.2. Design & Procedure

Recall that the different contexts in our study were designed to reflect the idea that contrast is not simply a binary notion, but rather that speakers' beliefs are gradient. We created four levels—*non-contradictory*, *weak*, *strong* and *presuppositional*—with the underlying assumption being that commitment would get increasingly stronger across these levels. The present task was conducted to test precisely this assumption, that is to directly measure **how strongly is A committed to "X Z-ed."** Thus, subjects who took part in this task only saw and rated Speaker A's part of the dialogue. The test was delivered via the web-based survey site Qualtrics. Participants sat in front of a computer screen and read a total of 24 contexts (A's part) pseudo-randomized among 24 fillers. On each trial, after reading A's context, they were asked to judge, on a scale from 1 to 7, how strongly is A committed to the fact that X Z-ed. So for instance, given context NO CONTR. in (12-a) above, participants were asked how strongly is Speaker A committed to the fact that "someone from the staff told Amy about her surprise party" (with 1 corresponding to extremely uncommitted and 7 to extremely committed). Here again, the procedure for English and French was exactly similar; French speakers provided judgments based on the question "*À quel point est-ce que A pense que X a Z?*"

2.3.3. Results

Results for both languages are reported in **Table 2**. Looking at the ratings descriptively, we indeed observe a strengthening trend across contexts. We see that context NO CONTR. is given the lowest commitment scores of all contexts, and that contexts STRONG, STRONG NAI and STRONG PRE.—which were designed to contain a stronger commitment of A to the prejacent proposition—are indeed being rated higher than contexts WEAK and WEAK NAI, which were meant to weakly commit A to the prejacent. Interestingly, we do not see a major difference between the strong and the presuppositional context.

Statistically, we fit a linear mixed effect model to the data for each language to determine whether participants' judgments varied depending on the fixed-effect predictor COMMITMENT. We were most interested in the following comparisons: comparing the context with no contradiction (NO CONTR.) to context with weak at-issue commitment (WEAK context), and comparing weak contexts (WEAK and WEAK NAI) to strong contexts (STRONG and STRONG NAI). We used sum-coding prior to analysis (i.e., -1/1) for each level in each comparison. The models included the maximal random effects structure justified by the data: random by-item intercepts, random by-participant intercepts and random slopes for COMMITMENT by item and participant. *P*-values were obtained by likelihood ratio test of the full model with the effect in question against the model without the effect in question. Concentrating on the comparison between our NO CONTR. and WEAK contexts, we found a significant effect of COMMITMENT both in English (β = 1.54, *SE* = 0.017, t = 3.31, $p < .001$), and French (β = 1.47, *SE* = 0.11, t = 3.56, $p < .001$). We also found a significant effect of COMMITMENT when comparing weak contexts to strong ones, both in English (β = 2.29, *SE* = 0.29, t = 5.24, $p < .001$), and French (β = 2.11, *SE* = 0.025, t = 4.98, $p < .001$). Overall, these results are welcome since they suggest that the contexts we designed *did* differ in the strength of A's commitment to a statement that B will contradict to the prejacent, and this for both languages. We can now turn to the main task, testing Hypothesis (iii).

TABLE 2 | Mean commitment judgments for pre-test 2.

	Mean ratings (subjects)		Mean ratings (objects)		Overall ratings	
	English	French	English	French	English	French
No contr.	2.2	2.3	2	2.1	2.1	2.2
Weak	3.6	3.8	3.9	4.1	3.8	4
Weak nai	2.7	3	2.6	3.5	2.7	3.2
Strong	6.1	6.4	6.1	6.1	6.1	6.25
Strong nai	5.5	5.8	5.3	5.8	5.4	5.8
Strong pre.	5.3	6	5.6	6.2	5.5	6.1

2.4. Main Task

2.4.1. Participants

We note that all participants in the main task were different from the participants who completed Task 1 and the Task 2.

For English: We recruited 64 participants on Amazon's Mechanical Turk with U.S. IP addresses (ages: 20–61; median: 36). They were paid $1 for their participation. Subjects who did not self-identify as native English speakers were not considered.

For French: We recruited 48 monolingual native speakers of French. All were given monetary compensation for their participation and were naive as to the goal of the experiment. Participants were from the regions of Pau, Toulouse and Albi in Southwestern France. Overall, 77% were undergraduate students, 17% graduate students, and 6% staff working at the university.

2.4.2. Design & Procedure

On each trial of this task, participants saw the whole dialogue, that is, A's background followed by Speaker B's comment (appearing either in cleft or canonical form). They were asked to judge the naturalness of B's sentence given A's on a seven-point Likert scale, with 1 corresponding to extremely unnatural and 7 to extremely natural.

We tested the effect of four factors on participants' ratings of cleft and canonical sentences: (i) EXISTENCE (based on measures collected in task 1), (ii) GRAMMATICAL FUNCTION (subject vs. object), (iii) AT-ISSUENESS, and (iv) CONTRARINESS. The factor CONTRARINESS was operationalized as the product of contradiction and strength of commitment (Contrariness = Contradiction $*$ Strength of Commitment). Contradiction, as mentioned in section 2.1.1, is either equal to 0 in the non-contradictory context (context NO CONTR. where Speaker B does not say anything that conflicts with what Speaker A says) or equal to 1 in the others. Consequently, items in the non-contradictory context had a contrariness value of 0. Items in contradictory contexts (contexts WEAK to STRONG PRE.) had a contrariness value equal to 1 (their contradiction value) * the value of Speaker A's commitment to the conflicting proposition, as measured in task 2.

If the data supports Hypothesis (iii), we expect to find that clefts are rated as more natural in the contexts where the level of contrariness is higher. We counterbalanced the experimental dialogues across 12 lists so that each participant judged a total 24 items (12 subjects and 12 objects). The order of the items was pseudo-randomized among 24 fillers.

2.4.3. Results

In the following, we begin by assessing our results descriptively, then we turn to the statistical analyses. Results combined for both sentence forms (clefts and SVO canonical sentences) and collapsed for grammatical function (subjects and objects) are illustrated in **Figure 1**, for English on the left panel and French on right panel. On **Figure 1**, red-colored markers represent clefts and black-colored markers represent canonical sentences. The cross-shaped markers on the y-axis indicate the naturalness ratings in the non-contradictory context (NO CONTR.). The circle-shaped markers indicate the ratings for the other contexts. Moreover, we note that the labels on the x-axis do not correspond to the number of our contexts, bur rather encode the contrariness values attributed to items in these contexts on a 7-point Likert scale, as per the results we gained in task 2 (discussed in section 2.3). Put simply, our x-axis represents the product of CONTRADICTION and COMMITMENT (as measured in task 2). Tables that include the mean naturalness ratings for each of our six contexts, per language and sentence form, can be found in **Appendix B**.

Inspecting the data for English, the figure reveals that the ratings for the cleft seem the most affected by CONTRARINESS, displaying the steepest increase across language and conditions (as illustrated by the upward trend in the position of the red dots). Indeed, clefts' ratings were the lowest of all in the NO CONTRNO CONTR. context (μ = 3.39), but increased as CONTRARINESS intensified (μ = 5.9 in STRONG PRE. context). The picture is quite different for canonical sentences: They were rated as very natural in the non-contradictory context (μ = 6.25), which should come to no surprise since in English, canonical sentences constitute an unmarked sentence form and are commonly used to answer an explicit wh-question. Interestingly, their felicity did not improve much with CONTRARINESS, but in fact slightly decreased (μ = 5.6). Despite this decrease though, canonical sentences were never judged infelicitous (in the sense of being below the midpoint of the 7 point scale), and were only slightly worse than clefts in the STRONG PRE. context.

Turning to French, we also observe an increase in clefts' naturalness as CONTRARINESS gets stronger, but to a much lower degree than in English. This is mainly due to the fact that French clefts are already rated fairly high in NO CONTR. context (μ = 4.56), as opposed to the English clefts (μ = 3.39), which is expected given that clefts are the most natural way to signal focus in the former language, especially with grammatical subjects (as argued by Lambrecht, 1994 among others, and empirically substantiated in Destruel, 2013; Féry, 2013). Similarly to English though, canonical sentences behave differently from clefts: While being rated highly in non-contradictory contexts (μ = 5.64), their naturalness does not improve as the level of CONTRARINESS rises (μ = 5.03). The first part of this result is interesting because it is at odds with many past accounts in the French literature that claim canonical sentences are highly disprefered in focus contexts (Lambrecht, 1994; Katz, 1997; Doetjes et al., 2004). What could be happening is that canonical sentences are rated as more felicitous in our study because they appear in written form, rather than in colloquial speech. We return to this point in the general discussion in section 3.

Now, we explore the data by grammatical function (subjects vs. objects), as illustrated in **Figure 2**, where ratings for canonical sentences appearing in the left panels and ratings for clefts appear in the right panels. The data for English are on the top two graphs; the data for French are at the bottom. On all plots, the red-colored markers represent the subject condition and black-colored markers represent the object condition. The cross-shaped markers represent the data for the non-contradictory context, and the circle-shaped markers represent the data for the contradictory contexts.

First, we concentrate on the right panels—the results for cleft sentences. Visual inspection of **Figure 2** reveals an asymmetry in clefts' ratings for French in the non-contradictory context (bottom right graph): Object clefts (black circles) appear clearly lower than subject clefts (red circles) (μ = 3.68 vs. μ = 5.43, respectively). This asymmetry relating to argument hierarchy is in line with the past literature and recent empirical evidence that suggest subject focus obligatorily induces a non-canonical structure while object focus only optionally does so since objects appear by default rightward, where prominence is assigned in French (Lambrecht, 2001; Destruel, 2016). We note that evidence for such an asymmetry is also provided cross-linguistically in languages such as Spanish (Buring and Gutierrez-Bravo, 2001), Northern Sotho (Zerbian, 2007), Georgian and Hungarian (Skopeteas and Fanselow, 2010b). This asymmetry is absent from our English data (see top right graph, μ = 3.26 and μ = 3.52 for objects vs. subjects, respectively), which is in line with the English results from an elicitation task reported in Skopeteas and Fanselow (2010b).

Second, looking at the data for canonical sentences, we see no such asymmetry in either of the two languages. Canonical sentences are rated equally high whether focus appears on the subject or the object, especially in the NO CONTR. context (English: μ = 6.3 for subjects and μ = 6.2 for objects; French: μ = 5.8 for subjects and μ = 6 for objects). Here as well, we note that the results for French are at odds with Lambrecht's claim that canonical sentences with lexical subjects is not the predominant pattern that surfaces in the spoken language (Lambrecht, 1987).

We conclude by reporting on the statistical analyses. We conducted mixed-effects linear regressions predicting clefts and canonical sentences' naturalness ratings in English and French from fixed effects of interest (i.e., grammatical function, at-issueness, existence and contrariness), and the following random-effect structure: random intercepts and slopes for the fixed effects of interest, and their interaction when relevant, per

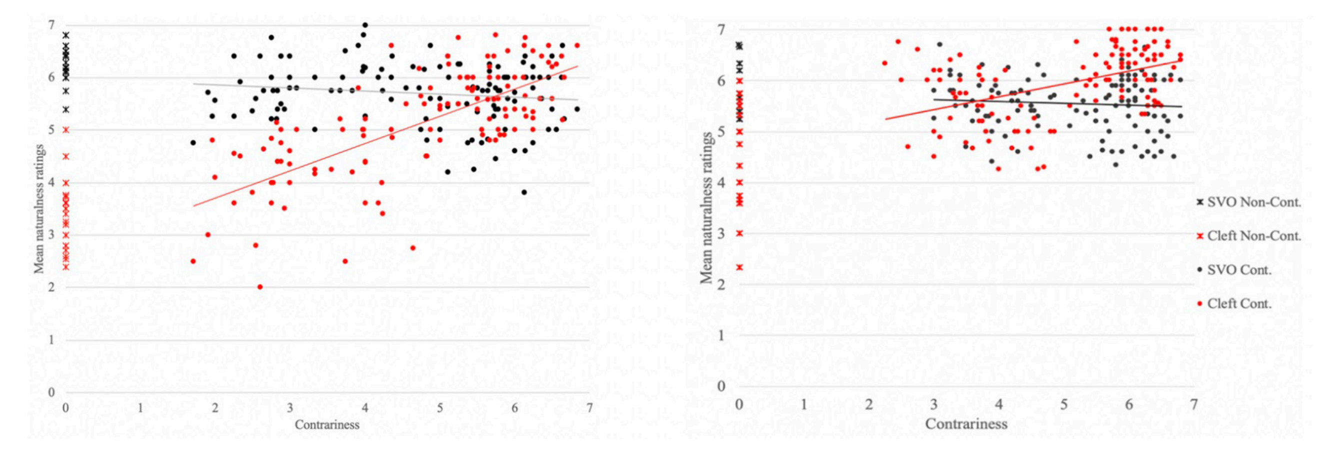


FIGURE 1 | Naturalness ratings for English **(Left)** and French **(Right)**.

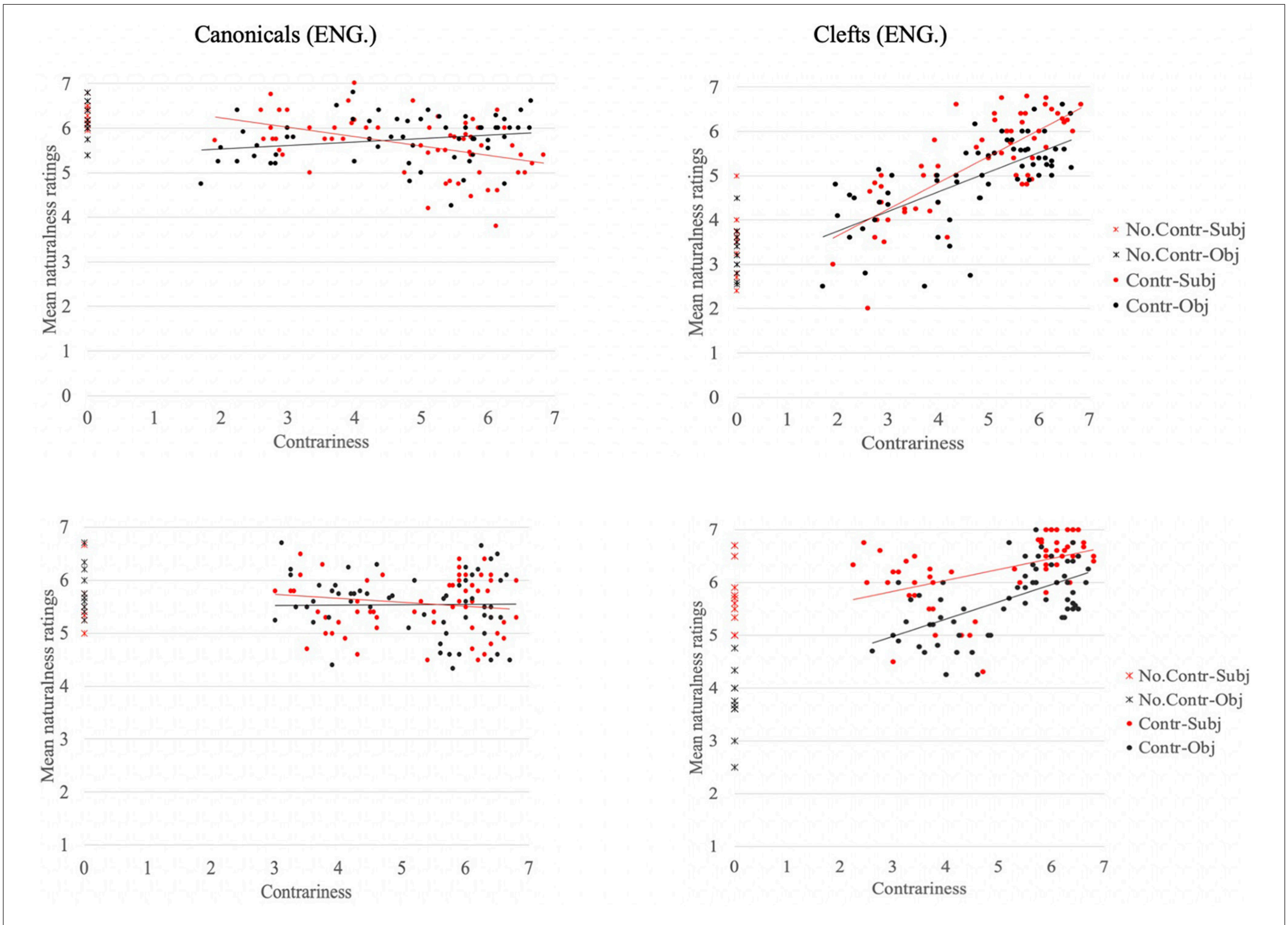


FIGURE 2 | Naturalness ratings per grammatical function, for canonical sentences **(Left)** and clefts **(Right)** in English **(Top)** and French **(Bottom)**.

participant and item. When the maximal models did not converge with the maximal random effects structure, they were re-conducted with the next maximal random effects structure until convergence was achieved. All fixed effects were centered before entering the analysis. To assess whether inclusion of a given factor significantly improved the fit of the overall model, likelihood-ratio tests were performed that compared two minimally different models, one with the fixed effects factor in question and one without, while keeping the random effects structure identical (Barr et al., 2013). The full model had the following structure: (Ratings $\sim$ 1 + At-Iss * GramFunct * Exist * Contrariness + Maximal RES). In the following, we report on estimates, standard errors, and t-values for all models (with any t-value exceeding 1.96 considered statistically significant with $p < 0.05$), as well as the χ^2 and P-values from the likelihood-ratio tests. Results were obtained using the *lmer* function of the *lme4* package (GPL-2|GPL-3, v.1.1-13; Bates et al., 2015) in the R environment (GPL-2|GPL-3, v.3.3.3; R Core Team, 2017).

We split the data prior to analysis, first looking at ratings for clefts in English. Within this data set, we found no effect of AT-ISSUENESS (β = -0.16, SE = 0.11, t = -1.44), suggesting that when clefts were used to signal an unexpected discourse move (i.e., to signal contrast on an element that was part of the non-at-issue content of A's speech) they were not drastically better than when commenting on an at-issue part of discourse. There was also no effect of GRAMMATICAL FUNCTION (β = 0.31, SE = 0.09, t = 1.13), such that the ratings for subject and object clefts were not significantly different. There was, however, an effect of EXISTENCE (β = 0.52, SE = 0.08, t = 6.55), suggesting that clefts were rated significantly better in contexts where the existence of the element to be contrasted is assumed. Of the three nested models, the one that gave the best fit to the data was the model that simply included the factor EXISTENCE (χ^2 = 9.72, $p < 0.01$). Of most interest to us, the factor that had the largest effect on clefts' ratings was CONTRARINESS (β = 0.01, SE = 0.001, t = 11.06)—the model that included this factor gave a significantly better fit to the data compared to the model that did not (χ^2 = 77.85, $p < 0.01$). This supports our hypothesis (Hypothesis iii) that clefts' naturalness is affected by the degree to which a speaker is committed to a (false) claim.

The picture is similar for French clefts in that AT-ISSUENESS had no effect either (β = -0.01, SE = 0.08, t = -1.34), but EXISTENCE did (β = 0.64, SE = 0.05, t = 12.03). One notable difference is that there was an effect of GRAMMATICAL FUNCTION (β = 0.88, SE = 0.07, t = 11.84), suggesting that subject clefts were given significantly better ratings than objects clefts. This result is unsurprising given what we already mentioned; that clefts are argued to be the default strategy to signal subject focus in French. The factor CONTRARINESS, although to a lesser extent than in English, also had a significant effect in predicting clefts' naturalness (β = 0.008, SE = 0.001, t = 5.55); a model that included this factor gave a better fit to the data than a model without it (χ^2 = 41.18, $p < 0.01$).

Finally, we report on the naturalness rating for the data set of ratings for canonical sentences. In English, we only found an effect of CONTRARINESS (β = -0.092, SE = 0.017, t = -5.24); all other factors did not significantly affect the felicity of canonical sentences. In French, we found no effect of AT-ISSUENESS (β = 0.11, SE = 0.10, t = 1.06) or of EXISTENCE (β = 0.09, SE = 0.08, t = 1.02), but there were an effect of GRAMMATICAL FUNCTION (β = -0.20, SE = 0.09, t = -2.25) and of CONTRARINESS (β = -0.007, SE = 0.001, t = -4.92).

3. GENERAL DISCUSSION

Clefts have long been noted to be focus-marking devices, often expressing a more special type of focus, i.e., contrastive focus, as opposed to a "simpler" type of focus generally referred to as *informational* focus (É. Kiss, 1998). Yet, traditional definitions of contrast appear unable to fully predict when these structures are most felicitous. This observation constituted the core motivation for our studies—our goal being to explore the relationship between the rhetorical role of focal alternatives and the naturalness of clefts in French and English, as per the two research questions in (10) and (11). More specifically, the experiments were designed to test the idea that clefts incorporate a requirement that the ordinary meaning is contrary to a previously salient focal alternative, which we operationalized via the notion of *contrariness* (i.e., strength of commitment * contradiction).

In the following, we first summarize the main experimental results and how they speak to our research questions, then we turn to discussing the implications of our findings for accounts on the meaning of clefts, definitions of contrast, and theories of focus.

Regarding the first research question, the experiment provided evidence that, although the presence of a focal alternative in the discourse context does increase the naturalness of clefts, it does not suffice to explain when clefts are preferred. In fact, while controlling for other factors known to influence the acceptability of clefts, naturalness ratings were significantly impacted when a doxastic contrast was involved: clefts are better in contexts where they indicate that an utterance runs contrary to a doxastic commitment held by the hearer, and the results are consistent with there being a requirement for a salient contrary doxastic commitment, whether that of an addressee or some other individual. We also found that whether contrastive content was marked as being at-issue or not did not significantly affect clefts' naturalness. This suggests that metalinguistic expectations about how a contrary point of view is changing in the discourse are less relevant to the acceptability of clefts than are salient beliefs about the world.

Our second research question asked whether dependency on the status of alternatives differs between French and English. In considering this question, it is necessary first to tease apart what we take to be independent differences between the two languages. Specifically, we need to separate the effects of grammatical function from the effects of the status of alternatives. Our experiments showed, in agreement with past literature, that in French but not English there is an effect of grammatical function: whereas in French subjects are more naturally clefted than objects, this is not the case for English. Our statistical analysis shows that once we control for this cross-cutting factor, we can see that clefts in the two languages exhibit very similar dependencies on the status of alternatives. In both languages clefts are more natural when there is doxastic contrast.

Even though our study was designed primarily to examine the use of clefts, another way to look at the data is to examine what happens in comparison with canonical sentences. It is often thought that their use is correlated: Lambrecht (1994) has claimed that clefts in French are used when canonical sentences are infelicitous. We find qualified support for this hypothesis, and indeed the effects are found in both languages we studied. On the one hand, canonical sentences were never rated as being highly infelicitous in our study. This fact appears to partially undercut Lambrecht's claim, since he motivated it on the basis of judgment and observational data suggesting that canonical sentences in French with lexical (i.e., non-pronominal) subjects are infelicitous. To the extent that we can operationalize infelicity as corresponding to mean ratings in the lower half of our 7 point scale, this is not what we found. While the results on clefts in non-contrastive conditions showed that the French speakers in our sample were prepared to mark at least some sentence types as being infelicitous in some conditions, they never rated canonical sentences as infelicitous. Thus, if French speakers only used clefts when their canonical counterparts were strictly infelicitous, they would be predicted to never use clefts at all, or at least not in any of the conditions we tested. Nonetheless, we did find reduced acceptability for canonical sentences in some conditions, specifically for sentences in French in which the context might lead to an expectation of focus on the subject, and for canonical sentences in both French and English for which the context led to a high level of contrariness. It is precisely in these conditions that cleft sentences have their highest mean acceptability in our study. Hence there is, at the very least, a correlation: the less acceptable canonical sentences are in a given context, the more acceptable corresponding cleft sentences are in that same context. It is thus plausible that at least one of the factors motivating cleft use is dispreference for use of the canonical form, albeit that it would be far too strong to say that cleft sentences are used when the canonical counterpart is unavailable.

What are we to make of the fact that canonical sentences in both French and English were judged to be slightly, but significantly degraded in contexts imposing a high degree of contrariness? One hypothesis consistent with this result is that the grammar directly imposes a penalty on the use of canonical sentences in such contexts. However, here, the style of Lambrecht's analysis provides an alternative way to look at the data. Lambrecht's model is paradigmatic, i.e., based on the contention that language users consider competing forms, and that suitability of one form depends on the availability and appropriateness of competing forms. It is consistent with the data that while the canonical form is unmarked, and has no requirements on (non-)contrariness, the cleft construction is a marked form which is specifically used when the meaning is also marked, for example in terms of contrariness. Thus in these situations, following what Horn (1984) called the division of pragmatic labor, the marked form is expected to be used in the marked context, and the unmarked form is then pragmatically dispreferred in these contexts. This type of explanation of the observed degradedness of canonical sentences in some contexts provides broad support for a Lambrechtian approach, even if his specific claims appear overly strong. Of course, it is also compatible with our data that cleft sentences are unmarked, and involve no inherent, conventionalized contrariness preference, but that canonical sentences have a conventional preference for non-contrary contexts. This seems a *prima facie* implausible analysis, reflected in the fact that the linguistic convention of terming the SVO form in English and French "canonical" already suggests that it is the unmarked form. We merely note that our data does not mitigate strongly against such an analysis.

As discussed in the introduction, the past literature on the meaning of clefts has largely characterized clefts as having three meaning components, cited in (2). Furthermore, much work has concentrated on describing the nature of exhaustivity, arguing either that it is semantically encoded in the cleft itself (Atlas and Levinson, 1981; Percus, 1997; É. Kiss, 1998; Hedberg, 2013), or that it arises as a result of pragmatic reasoning on the discourse context (Horn, 1981). In general, it is often supposed that aspects of meaning which are "baked" into the conventional meaning of an expression should surface more robustly than aspects of meaning and use which are derived indirectly, and involve pragmatic reasoning. Based on this premise, prior experimental research (Byram-Washburn et al., 2013; Destruel, 2013) has suggested that exhaustivity is pragmatic. The pattern of data that we have reported on in the current paper might then also be taken to suggest that contrariness requirements are derived via some pragmatic process, since, our contrariness data resemble prior exhaustivity data in that we observed gradient differences in judgments across conditions, rather than clear categorical effects with sharp boundaries between felicitous and infelicitous uses of clefts. However, we must note here that absent more constraints on possible conventional theories and the way they relate to judgment data, such a conclusion would be premature.

To see how our data might in principle be modeled in terms of linguistic conventions, let us briefly describe one such model. Call a base grammar one in which there is a certain set of requirements on the epistemic attitude of a salient individual toward a contrary proposition to the cleft. For example, this might be a null requirement, with no contrariness needed at all, it might be the requirement that a salient individual thinks the contrary proposition is possible, or it might be the requirement that a salient individual is certain of the contrary proposition. Now suppose that our experimental subjects are uncertain as to the exact meaning of a cleft, each entertaining a mixture of base grammars as possible models of the meaning of a cleft, and attributing different probabilities to each base grammar. Imagine that a person—for whom each base grammar G_i is assigned a non-trivial probabilities p_i—is faced with an example which is grammatical according to grammars $G_1, ..., G_r$, and ungrammatical according to grammars $G_{r+1}...G_n$. Let us suppose that their judgment of the grammaticality will be proportional to $\Sigma_1^r p_i$. That is, we suppose that felicity of an example is proportional to the likelihood of the grammar being one which accepts that example. In that case, the more contrary the context for an example, the more positive will be the predicted felicity judgment, since a more contrary case is bound to satisfy strictly more grammars. Further, the model would allow variation across experimental subjects to be modeled in terms of them having different base grammar probability distributions. Such a model could account for our gradient data entirely in terms of conventionally stipulated, categorical contrariness requirements of clefts. Thus, while we make no claim to have resolved whether contrariness is pragmatic (in which case an explanation of the phenomenon would still be needed), or based on a conventional requirement for contrariness, what we can say is that accounts of the meaning of clefts which are restricted to only the three standard components of cleft meaning are insufficient, since these do not account for our data.

Our research also relates to discussions on the definition of contrast concerned with how to characterize the nature of the alternatives in the interpretation of contrastive focus (as opposed to plain focus, or "informational" focus following É. Kiss, 1998). As discussed in the introduction, the past literature has often identified three relevant ingredients to contrast, namely the size of the alternative set, the identifiability of its elements, and the exclusion requirement of the alternatives. Our study speaks to the role of these aspects in that our experimental design included a **non-contrastive** context, in which these aspects were absent (i.e., an alternative to the focused element was not explicitly mentioned, and therefore what was said about the contrastively focused element potentially held of its alternatives), and **contrastive** contexts, in which the size of the alternative set was restricted to one alternative, explicitly mentioned (thus identifiable), and for which the predicate did not hold. Although we found that clefts' naturalness ratings were significantly better in the latter contexts for both languages, French clefts were rated fairly high in the non-contrastive context. This suggests that the presence of a clearly identifiable alternative (set) is not required in this language—the pivot position does not seem influenced by the alternative type, while it is in English. Thus, the grammatical sensitivity to this particular aspect of contrast differs between French and English.

Our main finding, though, suggests that characterizing contrast solely in terms of contrast set size, element identifiability and the exclusion requirement is insufficient. We have shown that the notion of *contrariness* is also important and indeed better explains clefts' use-conditions, both in French and English. This is where we would like to relate our finding to an idea present in Repp (2016): To gain a precise understanding of the notion of contrast, one should not only consider the way in which alternatives are construed, but also the type of context in which two sentences or discourse segments appear. Put slightly differently, Repp claims that while the alternativeness of constituents has to do with the explicitness (or lack there of) of the alternative (set), another important element of contrast has to do with the type of *discourse relation* in which sentences are involved. While the basic ingredients of contrast are that there must be similarities and dissimilarities between two sentences, Repp also discusses the fact that additional aspects can come into play—e.g., a violation of expectation—that lead to having a different discourse relation between two segments *d1* and *d2*. Repp hypothesizes that three relations are most relevant to the notion of contrast, which she calls NON-CONTRASTIVE, OPPOSE and CORRECTION relations. Crucially, she argues that these three discourse relations correspond to increasingly stronger degrees of contrast, which stems from the idea that contrast should indeed be considered a gradable phenomenon (an idea already present in some prior work such as Molnár, 2006). For instance, Repp argues that two segments in a CORRECTION relation express contrast more strongly than two segments that stand in an OPPOSE relation. The core difference between the three relations lies in the type of contribution that *d1* and *d2* make to the discourse: while *d1* and *d2* cannot be simultaneously true in an correction relation but can in a non-contrastive one, while they make opposing contribution in an oppose relation.

How do these discourse relations relate to the present work? In our experiment, given Repp's definitions for each relation, our contrastive contexts all involve a CORRECTION relation between the discourse segment of Speaker A and B—i.e., a piece of information in A is rejected by B, thus the propositions associated with the two segments cannot be simultaneously true. Therefore, although it would be tempting to try and explain our data in terms of differing discourse relations, our stimuli all stand in one and the same relation of correction. It would also be reasonable to cast an explanation of our data in terms of clefts being inherently corrective rather than inherently contrary. Put slightly differently, our notion of *contrariness* could be seen as an implementation of the notion of *corrective focus*: see e.g., Gussenhoven (2008). However, even though our analysis is inspired by correctivity accounts as well as by Zimmermann's, three differences are worth noting. First and most importantly, we take *contrariness* to be a matter of degree, which is not how corrections are normally analyzed. Indeed, existing accounts of correctivity do not incorporate any notion of degree, whereby one correction is in some sense stronger than another. In Repp's account, for example, the correction relation either holds between discourse segments or fails to hold, with no in between. The degree of contrast is encoded across relations, not within one. In our experiment, we varied the degree of contrast within the relation of correction. Therefore we can say that extant models of correctivity could not account for our data, and such models would have to be augmented in some way that would allow corrections of weakly held beliefs to be differentiated from corrections of strongly held beliefs.

Second, it seems plausible to have instances of contrariness where the claim runs counter to expectation but there has been no explicit counter claim to correct. Consider the example in (15). To deal with this example in a correction-based theory would require some modification to allow for the possibility of correcting things that have not actually been said, for instance by accommodation. Although, we do not dispute that such accommodation will sometimes be needed, we believe this is stretching the notion of correction unreasonably. Moreover, to make it work in Repp's account, which is a discourse-relation based account, one would need accommodation of a contrary utterance. Because our contrariness account is based not on differences between what has been said, but on the difference between beliefs, to account for cases like (15), we require a different type of modification, namely accommodation of contrary belief. We recognize however, that this is a quite subtle difference between Repp's corrective account and our contrariness account; a difference as to whether they focus on what is believed vs. what is said, and that it might be hard to find examples that truly distinguish between the two.

(15) A: Who won the NBA dunk contest this year?
B: No way you'll believe this, but it just so happens that it was an unknown contender from Iowa called Louis D. Johnson who managed to get the most points, and on the

final dunk!
A: No fucking way!
B: Yes fucking way - you should have seen her alley-oop windmill off the back of a donkey? Johnson is incredible!!!

Third, Prince's informative clefts are a problem for corrective analyses. Consider for instance example (16):

(16) It was at the University of Iowa that Camille D. Johnson first managed to apply her deep knowledge of clefts in natural language to the world of particle physics, and, for the first time in human history, to split the atom entirely by the use of carefully targeted questions.

While there is no prior material being corrected here, it is not implausible that, in such cases, the claim is being presented for rhetorical effect as running counter to an expectation. Here again though, it is implausible that we could accommodate that someone had said something contrary to this, but it is quite plausible that we could accommodate a contrary belief or expectation[8].

Finally, our findings can be considered in the broader light of prior work on the function of prosody and other ways of marking information status. Much prior work on focus has emphasized properties that relate to the presence of some prior structure in discourse, for example the presence of a question, of an element of the same type as the target, or of a clause which exhibits structural parallelism. A different line of work was initiated by Pierrehumbert and Hirschberg (1990), who analyze various types of intonational contour in terms of speaker and hearer expectations. Our experiments and analysis imply that clefts have an intrinsically doxastic function. While the specific results we have obtained are not predicted by any prior model, they do suggest that the Pierrehumbert and Hirschberg approach is on right track for analyzing the marking of information status more generally.

Indeed, they are also in line with work suggesting that marking of speaker expectation is a central function of language, markers of such expectations sometimes being brought together in a (controversial) category of *miratives* (DeLancey, 1997). It is notable that several focus sensitive constructions have been taken to be mirative, including scalar additives like English *even* / French *même* and exclusives like English *only* / French *seulement* (see e.g., Beaver and Clark, 2009). Recently, Cruschina (2012) discusses the relationship between contrast and *focus fronting* (i.e., movement of the focus constituent to the left-periphery of the sentence), arguing that different subtypes of focus are relevant for the realization and interpretation of this syntactic movement. For instance, the author shows that while most Romance languages employ focus fronting as a strategy to signal the most explicit case of contrast, namely *correction*, they also resort to this strategy to encode mirative focus, that is new information that is particularly surprising or unexpected to the hearer. In the same vein, Trotzke (2017) provides empirical evidence for German. Results from an acceptability judgment task suggest that focus fronting in this language is also more commonly associated with a mirative interpretation rather than either a corrective or a contrastive interpretation. Finally, Bianchi et al. (2016) find that the intonational patterns associated with fronted constituents in the mirative condition differs from those found in the correction condition, thus positing that mirative focus is indeed grammatically distinct from corrective focus. The authors go on discussing the nature (or status) of this mirative interpretative effect, analyzing it as a conventional implicature (in the sense of Potts, 2005). Going back to the construction of interest in this paper, the fact that clefts, which help mark focus, turn out to have a function related to speaker expectation is of a piece with the fact that some focus sensitive constructions have previously been identified as mirative. Although we are not currently in a position to make any strong claims about the nature of the contrariness requirement we posit for clefts, we acknowledge that the analysis proposed by Bianchi et al. (2016) for the mirative effects associated with focus fronting might be extendable to clefts. Given the amount of work on the nature of exhaustivity in clefts, which is analyzed as an implicature by some (see e.g., Horn, 1981; Destruel, 2013), one line of research worth pursuing would be to directly compare the strength of the contrariness effect with the effects of classic inferences failing such as exhaustivity.

While our data answers the main questions we set out with, it is also suggestive of new questions. First, we might ask whether the judgment effects we have observed would be mirrored in usage data, e.g., in terms of the frequencies of canonical sentences and cleft sentences in more or less contrary contexts. Indeed, although rating scales tend to provide stable, replicable and transparent pieces of data (Tonhauser and Matthewson, unpublished manuscript), one limitation concerns the possible

[8]We contend that to date, no theory that requires an explicit discourse antecedent can account for informative clefts without some additional mechanism of accommodation. However, in augmenting the theory in this way, corrective analyses face a conceptual problem which is absent from the contrariness analysis. The term *correction* is standardly used in linguistics to refer to a speech act used to correct another speech act, or to a rhetorical relation holding between pairs of speech acts. To the extent that our contrariness proposal assumes anaphoricity of the cleft to a discourse antecedent, it faces the same problem. However, the central notion of our contrariness proposal is not disagreement with a speech act, but disagreement with an attitude. To be clearer, there are two distinct requirements in the contrariness theory. If Repp or some other supporter of a correction-based account was to understand the notion of correction more liberally, such that it could include correction of implicit assumptions, then indeed this data would not divide between correction and contrariness accounts. In that case, data like (16) is showing what sort of correction-based account is needed, rather than showing that the notion of correction should be entirely dispensed with. Yet, we believe examples such as (i) might distinguish between the two proposals.

(i) Emilie (looking disappointed in David): You ate the rest of my cake, didn't you?
David (wiping crumbs from mouth and looking guilty): Uhhh, ooops. Yeah, ok, it WAS me who ate it, but I thought you were done, and it looked way too good to waste!

In this type of case, there is clearly no correction since the speaker is agreeing with the suggestion from the speaker's slightly biased question. There is contrariness however, but it is contrariness to norms or expectations about proper behavior, not contrariness with respect to what Emilie said or believed at utterance time. Therefore, our idea is that contrariness is a much more natural notion than correction here. One does not "correct" norms or expectations by showing a counter-instance.

variation in participants' interpretation of the provided Likert scale, and therefore their resulting use of the scale to provide their judgments. In the present studies, we chose to label each point on the Likert scale rather than only the two end points in order to limit variation as much as possible. Another potential limitation concerns the fact that the language data we are examining involve quite subtle judgments, which might explain the gradience we observe in our results. Given these potential limitations, a corpus investigation would be a welcome addition but such an investigation is not necessarily straightforward and easy to implement, as it would require the operationalization of the notion of contrariness in naturally occurring data. This would certainly be a challenge with a purely automatic methodology for identifying examples in corpora, but perhaps is not beyond what might be achieved using a combination of computational methods for retrieving naturally occurring clefts in context, and human annotation for assessing the degree of contrariness (or, for that matter, correctivity, if this could be assessed as a matter of degree).

Second, for those who accept the premise that gradient data of the sort we see in this experiment is suggestive of a pragmatic rather than a semantic account, what would be the underlying pragmatic explanation? That is, how might one derive from standard assumptions about the meaning of clefts and standard pragmatic principles the fact that clefts are more felicitous as contrariness increases? Finally, given that we have established that in some way clefts are used to mark differences in expectation, how might they be fitted into a more general theory of mirativity, i.e., of how expectation is signaled in human language?

4. CONCLUSIONS

The goal of the present paper was to test prior hypotheses concerning clefts' standard components of meaning. We hypothesized that the mere presence of an antecedent in discourse which the clefted element would pick up and comment on (i.e., simple contrast) would not suffice to fully explain the felicity pattern of English *it*-clefts. Instead, we set out to test the hypothesis that something more refined is needed, namely a notion of contrast that includes a conflict between interlocutors' expectations. We adapted Zimmermann's notion of contrast, which relates to how strongly the addressee believes the contrary, and experimentally operationalized it. Our data suggests that contrariness does indeed play an important role in helping speakers choose between cleft and canonical forms: the more strongly an interlocutor appears committed to a false proposition, the better it is to repudiate them with a cleft as opposed to using canonical word order, and this effect is visible over and above other factors that distinguish the distribution of clefts in French and English.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the IRB at the University of Texas with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the IRB.

AUTHOR CONTRIBUTIONS

ED designed and ran the experiments, collected the data, analyzed the data, and wrote the first draft of the paper. DB helped with ideas in the discussion section and wrote a second draft of that section. EC helped with providing comments on the overall paper, and reworded part of the introduction.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01400/full#supplementary-material

REFERENCES

Atlas, J., and Levinson, S. (1981). "It-clefts, informativeness, and logical form: radical pragmatics (revised standard version)," in *Radical Pragmatics*, ed P. Cole (New York, NY: Academic Press), 1–61.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beaver, D. I., and Clark, B. Z. (2009). *Sense and Sensitivity: How Focus Determines Meaning*, Vol. 12. Oxford: John Wiley & Sons.

Bianchi, V., Bocci, G., and Cruschina, S. (2016). Focus fronting, unexpectedness, and evaluative implicatures. *Semant. Pragmat.* 9, 1–54. doi: 10.3765/sp.9.3

Buring, D., and Gutierrez-Bravo, D. R. (2001). "Focus-related word order variation without the nsr: a prosody-based crosslinguistic analysis," in *Syntax at Santa Cruz 3* (Santa Cruz, CA), 41–58.

Büring, D., and Kriz, M. (2013). It's that, and that's it! exhaustivity and homogeneity presupposition in clefts (and definites). *Semant. Pragmat*. 6, 1–29. doi: 10.3765/sp.6.6

Byram-Washburn, M., Kaiser, E., and Zubizarreta, M. L. (2013). *The English It-Cleft: No Need to Get Exhausted.* Boston, MA: Poster, Linguistic Society of America (LSA).

Carter-Thomas, S. (2009). "The French *c'est*-cleft: functional and formal motivations," in *La linguistique Systémique Fontionnelle et la Langue Française*, eds D. Banks, S. Eason, and J. Ormrod (Paris: L'Harmattan), 127–156.

Cruschina, S. (2012). *Discourse-Related Features and Functional Projections.* Oxford: Oxford University Press.

Declerck, R. (1984). The pragmatics of it-clefts and WH-clefts. *Lingua* 64, 251–289. doi: 10.1016/0024-3841(84)90065-2

DeLancey, S. (1997). Mirativity: the grammatical marking of unexpected information. *Linguist. Typol.* 1, 33–52. doi: 10.1515/lity.1997.1.1.33

Delin, J. (1991). "Towards a model for generating cleft sentences," in *Pragmatics at Issue. Selected Papers of the International Pragmatics Conference* (Amsterdam; Philadelphia, PA: John Benjamins), 113–132.

Destruel, E. (2013). *An Empirical Study on the Meaning and Use of the French* c'est-*Cleft*. PhD thesis, University of Texas at Austin.

Destruel, E. (2016). Focus marking asymmetries in colloquial and standard french: a stochastic optimality-theoretic account. *J. French Lang. Stud.* 26, 299–326. doi: 10.1017/S0959269515000265

Destruel, E. and DeVeaugh-Geiss, J. (2018). On the interpretation and processing of exhaustivity: evidence of variation in English and French clefts. *J. Pragm.* 138, 1–16. doi: 10.1016/j.pragma.2018.09.009

Destruel, E., and Velleman, L. (2014). "Refining contrast: empirical evidence from the English *it*-cleft," in *Empirical Issues in Syntax and Semantics 10*, ed C. Piñón (CSSP), 197–214.

Destruel, E., Velleman, L., Onea, E., Bumford, D., Xue, J., and Beaver, D. (2015). "A cross-linguistic study of the non-at-issueness of exhaustive inferences," in *Experimental Perspectives on Presuppositions*, ed F. Schwarz (New York, NY; Dordrecht; London: Springer), 135–156.

Doetjes, J., Rebuschi, G., and Rialland, A. (2004). "Cleft sentences," in *Handbook of French Semantics*, eds F. Corblin and H. de Swart (Stanford, CA: CSLI), 529–552.

Féry, C. (2013). Focus as alignment. *Nat. Lang. Linguist. Theory* 31, 683–734. doi: 10.1007/s11049-013-9195-7

Gussenhoven, C. (2008). "Types of focus in english," in *Topic and Focus,* eds C. Lee, M. Gordon, and D. Büring (Heidelberg; New York, NY; London: Springer), 83–100.

Halvorsen, P.-K. (1978). *The syntax and semantics of cleft constructions*. (Ph.D. thesis), Austin, TX: University of Texas at Austin.

Hamlaoui, F. (2008). "Focus, contrast, and the syntax-phonology interface: the case of French cleft-sentences," in *Current Issues in Unity and Diversity of Languages: Collection of the Papers Selected From the 18th International Congress of Linguists* (Seoul: Linguistic Society of Korea).

Harries-Delisle, H. (1978). "Contrastive emphasis and cleft sentences," in *Universals of Human Language*, ed J. H. Greenberg (Stanford: Stanford University Press), 419–486.

Hedberg, N. (1990). *The discourse function of cleft sentences in English*. (Ph.D. thesis), Minneapolis, MI: University of Minnesota.

Hedberg, N. (2013). "Multiple focus and cleft sentences," in *Cleft Structures*, eds K. Hartmann and T. Veenstra (Amsterdam: John Benjamins Publishing CompanyJohn Benjamins Publishing Company), 227–250.

Horn, L. (1984). "Towards a new taxonomy of pragmatic inference: Q-based and r-based implicatures," in *Meaning, Form and Use in Context*, ed D. Schriffin (Washington, DC: Georgetown University Press), 11–42.

Horn, L. R. (1981). "Exhaustiveness and the semantics of clefts," in *Proceedings of NELS 11* (Boston, MA), 125–142.

Jespersen, O. (1927). *A Modern English Grammar on Historical Principles: Part VII: Syntax. (Completed and Edited by Niels Haislund)*. Heidelberg: Carl Winters. (Reprinted 1961 Allen & Unwin, London and Ejnar Munksgaard, Copenhagen).

Katz, J., and Selkirk, E. (2011). Contrastive focus vs. discourse-new: evidence from prosodic prominence in english. *Language* 4, 771–816. doi: 10.1353/lan.2011.0076

Katz, S. L. (1997). *The syntactic and pragmatic properties of the* c'est-*cleft construction*. (Ph.D. thesis), University of Texas at Austin, Austin, TX.

Kenesei, I. (2006). "Focus as identification," in *The Architecture of Focus*, eds V. Molnár and S. Winkler (Berlin: Mouton de Gruyter), 143–159.

Kiss, K. É. (1998). Identificational focus versus information focus. *Language* 74, 245–273. doi: 10.1353/lan.1998.0211

Krifka, M. (2008). Basic notions of information structure. *Acta Linguist. Hung.* 55, 243–276. doi: 10.1556/ALing.55.2008.3-4.2

Lambrecht, K. (1987). "On the status of SVO sentences in French discourse," in *Coherence and Grounding in Discourse*, ed R. S. Tomlin (Amsterdam: John Benjamins), 217–262.

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus and Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics* 39, 463–516. doi: 10.1515/ling.2001.021

Lopez, L. (2009). *A Derivational Syntax of Information Structure*. Oxford: Oxford University Press.

Molnár, V. (2002). Contrast from a contrastive perspective. *Lang. Comput.* 39, 147–162. doi: 10.1163/9789004334250_010

Molnár, V. (2006). "On different kinds of contrast," in *The Architecture of Focus,* eds V. Molnar and S. Winkler (Berlin: Walter de Gruyter), 197–234.

Patten, A. (2012). *The English It-cleft: A Constructional Account and a Diachronic Investigation*. Amsterdam: De Gruyter.

Percus, O. (1997). "Prying open the cleft," in *Proceedings of the 27th Annual Meeting of the North-East Linguistics Society (NELS)*. ed K. Kusumoto (Amherst: GLSA).

Pierrehumbert, J., and Hirschberg, J. (1990). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication,* eds P. Cohen, J. Morgan, and M. Pollack (Cambridge, MA: MIT Press), 271–312.

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.

Prince, E. F. (1978). A comparison of wh-clefts and it-clefts in discourse. *Language* 54, 883–906. doi: 10.2307/413238

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Repp, S. (2016). "Contrast: dissecting an elusive information-structural notion and its role in grammar," in *Handbook of Information Structure*, eds C. Féry and S. Ishihara (Oxford: Oxford University Press), 270–289.

Rooth, M. (1985). *Association with Focus*. Amherst, MA: Graduate Linguistics Student Association, UMass.

Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semant.* 1, 75–116. doi: 10.1007/BF02342617

Sarnicola, R. (1988). It-clefts and WH-clefts: two awkward sentence types. *J. Linguist.* 24, 343–379. doi: 10.1017/S0022226700011828

Selkirk, E. (2008). Contrastive focus, givenness and the unmarked status of "discourse-new". *Acta Linguist. Hungar.* 55, 331–346. doi: 10.1556/ALing.55.2008.3-4.8

Skopeteas, S., and Fanselow, G. (2010a). Focus in Georgian and the expression of contrast. *Lingua* 120, 1370–1391. doi: 10.1016/j.lingua.2008.10.012

Skopeteas, S., and Fanselow, G. (2010b). "Focus types and argument asymmetries: a cross-linguistic study in language production," in *Contrastive information structure*, eds. C. Breul and E. Gobbel (Amsterdam; Philadelphia, PA: John Benjamins Publishing Company), 169–197.

Thoma, S. (2009). "St'àt'imcets independent pronouns-the invisible cleft," in *Proceedings of the 23rd Northwest Linguistics Conference. Working papers of the Linguistics Circle of the University of Victoria* (Seattle, WA), 109–123.

Trotzke, A. (2017). Mirative fronting in german: experimental evidence. *Rev. Cogn. Linguist.* 15, 460–488. doi: 10.1075/rcl.15.2.07tro

Umbach, C. (2004). On the notion of contrast in information structure and discourse structure. *J. Semant.* 21, 155–175. doi: 10.1093/jos/21.2.155

Vallduvi, E., and Vilkuna, M. (1998). "On rheme and kontrast," in *The Limits of Syntax*, eds P. Culicover and L. McNally (New York, NY: Academic Press), 79–108.

Velleman, D., Beaver, D., Onea, E., Bumford, D., Destruel, E., and Coppock, E. (2012). "It-clefts are IT (Inquiry Terminating) constructions," in *Proceedings of Semantics and Linguistic Theory (SALT) 22* (Chicago, IL: eLanguage), 441–460.

Wedgewood, D. (2007). "Identifying inferences in focus," in *On Information Structure, Meaning and Form*, eds K. Schwabe and S. Winkler (Amsterdam: John Benjamins), 207–227.

Yasavul, M. (2013). "Two kinds of focus constructions in k'iche," in *Proceedings of the 23rd Semantics And Linguistic Theory Conference, UCSC, CA* (Chicago, IL).

Zerbian, S. (2007). "Subject/object-asymmetry in Northern Sotho," in *Information Structure and the Architecture of Grammar: A Typological Perspective,* eds K. Schwabe and S. Winkler (Amsterdam: John Benjamins Publishing Company), 323–345.

Zimmermann, M. (2008). "Contrastive focus and emphasis," in *Acta Linguistica Hungarica is a Quarterly Hungarian Peer-reviewed Academic Journal in the Field of Linguistics*, ed F. Kiefer (Akadémiai Kiadó), 347–360.

Zimmermann, M. (2011). The expression of focus in west chadic: variation and uniformity in and across languages. *Linguistics* 49, 1163–1213. doi: 10.1515/ling.2011.032

Zimmermann, M., and Onea, E. (2011). Focus marking and focus realization. *Lingua* 121, 1651–1670. doi: 10.1016/j.lingua.2011.06.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 July 2019
doi: 10.3389/fpsyg.2019.01461

How Do Addressees Exploit Conventionalizations? From a Negative Reference to an *ad hoc* Implicature

Edmundo Kronmüller[1] and Ira Noveck[2]*

[1]Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile, [2]Institut des Sciences Cognitives – Marc Jeannerod (UMR 5304), Lyon, France

A negative reference, such as "not the sculpture" (where *the sculpture* is a name the speaker had only just invented to describe an unconventional-looking object and where the negation is saying that she does not currently desire that object), seems like a perilous and linguistically underdetermined way to point to another object, especially when there are three objects to choose from. To succeed, it obliges listeners to rely on contextual elements to determine which object the speaker has in mind. Prior work has shown that pragmatic inference-making plays a crucial role in such an interpretation process. When a negative reference leaves two candidate objects to choose from, listeners avoid an object that had been previously named, preferring instead an unconventional-looking object that had remained unnamed (Kronmüller et al., 2017). In the present study, we build over these findings by maintaining our focus on the two remaining objects (what we call the *second* and *third* objects) as we systematically vary two features. With respect to the second object – which is always unconventional looking – we vary whether or not it has been given a name. With respect to the third object – which is never named – we vary whether it is unconventional or conventional looking (for the latter, imagine an object that clearly resembles a bicycle). As revealed by selection patterns and eye-movements in a visual-world eye-tracking paradigm, we replicate our previous findings that show that participants choose randomly when both of the remaining objects are unconventional looking and unnamed and that they opt reliably in favor of the most nondescript (the unnamed unconventional looking) object when the second object is named. We show further that (unnamed) conventional-looking objects provide similar outcomes when juxtaposed with an unnamed unconventional object (participants prefer the most non-descript as opposed to the conventional-looking object). Nevertheless, effects emerging from the conventional (unnamed) case are not as strong as those found with respect to those reported when an unconventional object is named. In describing participants' choices in the non-random cases, we propose that addressees rely on the construction of an *ad hoc* implicature that takes into account which object can be eliminated from consideration, given that the speaker did not explicitly name it.

Keywords: pragmatics, negation, reference, conventions, *ad hoc* implicature, eye-tracking

OPEN ACCESS

Edited by:
Katharina Spalek, Humboldt University of Berlin, Germany

Reviewed by:
William Levine, University of Arkansas, United States
Alexia Galati, University of North Carolina at Charlotte, United States

****Correspondence:***
Edmundo Kronmüller
ekr@uc.cl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *31 December 2018*
Accepted: *07 June 2019*
Published: *03 July 2019*

Citation:
Kronmüller E and Noveck I (2019) How Do Addressees Exploit Conventionalizations? From a Negative Reference to an ad hoc Implicature. Front. Psychol. 10:1461. doi: 10.3389/fpsyg.2019.01461

INTRODUCTION

Imagine you are asked to assist a jewelry shop owner as she prepares a window display of newly arrived, hard-to-describe brooches. While you are at the storefront poised to display a brooch, the owner is sitting behind a computer screen while looking at two brooches at a time in order to make thoughtful comparisons. Over the course of her deliberation, the owner refers to the brooches by giving approximate names for them (e.g., the *ballerina*, the *insect*, etc.). You now pull out a box-set of three other hard-to-describe brooches from an up-and-coming jewelry-designer as the owner, again, ponders two at a time. While sharing her impressions, she soon refers to one of the three in the box as "the one that looks like a modern sculpture," which you can now identify. While pondering over which one of the three she wants to display, she finally shouts out what we call a negative reference (Kronmüller et al., 2017) – "not the sculpture" – because she actually wants the other one on her screen. Which one is she referring to? Prior experimental work, which captures just such a situation in more austere conditions, has shown that when the speaker employs "not" in such a context, the addressee (the participant) will randomly choose from the remaining two under consideration (given that there is no possibility for the listener to ask the speaker which one). This makes sense given that, from the addressee's perspective, there remain two possibilities out of three. Interestingly, the work on negative reference further shows that when two of the three objects have been given a name (to return to the scenario above, imagine one brooch has been coined the *sculpture* and a second one the *bench*), listeners will typically eliminate the second, named object from consideration as well (and reliably so, at a rate of about 80%) while ultimately pointing to the only remaining unnamed object (see Kronmüller et al., 2017). See **Figure 1** for a representation of these two, critical (*baseline* and *second-object-is-named*) conditions. Such phenomena reveal that addressees assume that an interlocutor will use an agreed-upon name when it is available. The current work extends the prior work by employing this confirmed paradigm to investigate the case in which an unnamed object – one that could have a conventional name (e.g., imagine an iconic representation of a bicycle or a gamepad) – similarly determines performance on such a negative reference task.

In the remainder of the section "Introduction," we will do three things. First, we will describe how the task reported in this study came into being, and we will present its features in greater detail so as to elucidate how prior studies have been useful in making claims about reference. Second, we will show how key results from this class of experiments can be viewed as *ad hoc* implicatures, as introduced in the scalar inference literature (Papafragou and Tantalou, 2004; Barner et al., 2011; Stiller et al., 2015). Finally, we will describe how we modified the task for the purposes of the present study in order to investigate the role of conventionality with respect to this pragmatic inference.

Background

This particular paradigm actually evolved without negations to address a debate that aimed to answer the following questions: does a listener incorporate who named an (unusual looking) object when two different speakers are doing the naming independently, or does the listener treat the object name as author-free? Consider the case of a voice (heard over headphones) that helpfully identifies an unusual looking object for the participant by calling it "the thing that looks like a modern sculpture" and eventually just "the sculpture." When a new voice also calls the object *the sculpture*, do participants' looks go directly to the object (on a screen) without hesitation? Are participants surprised when the second voice (these tasks are usually carried out with distinctive male and female voices) comes up with a new name for an object? Prior work had led to multiple eye-tracking experiments with straightforward affirmative references (Barr and Keysar, 2002; Metzing and Brennan, 2003; Kronmüller and Barr, 2007; Brown-Schmidt, 2009), all designed to capture how immediately a person looks at and clicks on a previously named object when that object is referred to by a new speaker (this paradigm also detects

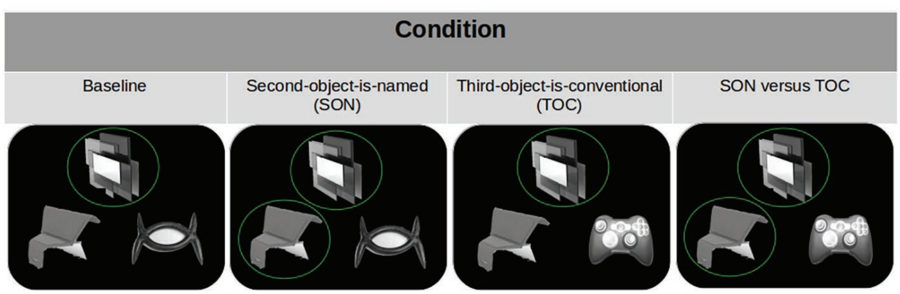


FIGURE 1 | An example of a test trial screen in each condition whose final instruction is "not the sculpture." These are, from left to right, the baseline, the second-object-named (SON), the third-object-conventional (TOC), and the SON-vs.-TOC conditions. The upper object in each is the negated reference that has been heretofore mentioned twice as "the sculpture." A green circle indicates that the object (an unconventional one) has been given a name (circles do not appear on the participant's screen). In these examples, the object on the lower left hand side is the second (Y) object that has been called "the bench" (twice until now) in the second-object-named (SON) condition as well as in the SON-vs.-TOC condition. The object on the lower right hand side is the third (Z) object; this object can be rendered conventional and readily identifiable (as a "gamepad") in the TOC and SON-vs.-TOC conditions.

how much confusion is produced when the same speaker changes an object's previously given name). In a meta-analysis of this work, Kronmüller and Barr (2015) showed that participants find objects without fully paying heed to the source of the coinage; rather, they pay attention to the fact that an object has a name attached to it.

The negative reference study in Kronmüller et al. (2017), described above, was built to further explore the timing involved with respect to the source of coinage (by including two speakers) when resolving reference. In the present work, we aim to exploit this paradigm's reliable findings when faced with a *solitary* speaker. Our goal is more fundamental, which is to determine whether a conventional-looking object that remains unnamed holds as much sway as an unconventional-looking object that has been referred to with an agreed-upon name. In this way, we can explore how two different kinds of conventions (one being the invented names briefly shared by interlocutors and the other being conventional representations that are presumably shared by a language community) are perceived by participants in a single task.

Before turning to the current experiment, let us carefully review the details of Kronmüller et al.'s (2017) original negative reference paradigm in order to fully appreciate its pragmatic features. As indicated, a participant (the addressee) is viewing three unusual-looking objects – let us call them X, Y, and Z – after being told that the speaker is viewing two objects and while assuming that the participant is seeing the same two. In the critical, experimental (what we will call the *second-object-is-named*) condition, the speaker has provided two of the three objects (X and Y) with names. What prompts a significant majority of participants to choose the least-familiar object, Z, as the intended object of "not the X" in this scenario? To start, when a participant hears "not the X," it reveals that one of the speaker's observable objects indeed includes the previously-named X. It also prompts the listener to create an *ad hoc* category of two objects (for seminal descriptions of *ad hoc* categories, see Barsalou, 1983), containing a second named object (Y) and an unnamed one (Z). The question is which of the two is paired with X (see **Table 1** below). A listener could arguably conclude that it is more unlikely that the speaker would refer to Y as "not the X" because the speaker had previously referred to Y with a name, just as she had used X in referring to it with a name. This makes it more likely that X is paired with Z. The supposition that the speaker had Z in mind is thus optimal for resolving the reference "not the X."

Prior experimental pragmatic investigations into *ad hoc* pragmatic inference emerged with respect to children's production of scalar implicatures. Scalar implicature refers to cases in which a relatively weak expression, such as *Some of the cats are black*, is thought to imply the rejection of a more informative and unsaid one (such as *All of the cats are black*) to yield the implicature *Not all of the cats are black*. Children are widely known to be less likely than adults to make this pragmatic inference (e.g., see Noveck, 2001, 2018; Katsos et al., 2016). As the reliability of the developmental effect grew and as explanations generally relied on participants' knowledge of linguistic scales related to lexical terms, i.e., how it relies on recognizing that

TABLE 1 | A representation, from the listener's perspective, of the two possible pairings that the speaker is viewing when saying "Not the X," in the Kronmüller et al. (2017) paradigm.

Possibility 1		OR	Possibility 2	
X	Y		X	Z

All entails *Some*,[1] Papafragou and Tantalou (2004) investigated cases in which children can equally or more reliably exploit *ad hoc* categories in context (*Did the cow wrap the gifts? He wrapped the parrot*); these *particularized* cases (as opposed to *generalized* cases) bypass concerns about linguistic competencies, such as knowledge about and the application of lexical scales (see Grice, 1989). For an illustrative example of *ad hoc* implicature development, consider work from Stiller et al. (2015), who investigate cases in which 2- to 5-year-old children as well as adults are shown three smiley faces – one classic smiley face, a second smiley face wearing glasses, and a third wearing glasses and sporting a hat. When participants are told "My friend wears glasses," it is at around three-and-a-half years of age that children reliably point to the smiley face wearing glasses only, even though there are two smiley faces with glasses to consider. This more precise reading of the utterance (to mean *the friend is wearing glasses but no hat*) is wholly contextual and occurs on the fly based on (1) the *ad hoc* category of three presented smiley-faces (i.e., without concerns about lexical scales), and while (2) inferring that the speaker would have said "my friend is wearing a hat" if that was indeed the friend the speaker had in mind. In fact, adult-like performance appears to emerge earlier among children in experimental situations that call on *ad hoc* implicatures when compared to those that rely on knowledge about linguistic terms and scales.

As should be clear, *ad hoc* implicatures similarly come into play in the negative reference task, whose contextual cues create an *ad hoc* category from which one can make more precise interpretations; in this case, the process begins with a negative reference. When a speaker says "not the X," leaving two objects to be considered as a target, a listener's interpretation (which of the two did the speaker have in mind?) depends on what he knows about prior references. In what is essentially the control condition (what we will refer to later as *the baseline*), where X is the only object with a name, listeners have no reason to favor one of the unknown objects as the speaker's referent over the other; *Not the X* ought to lead to random responding among the options Y and Z, as has been reported. When one of the remaining objects (Y) has also been given a name, however, listeners are more likely to exclude it from consideration as a partner for X and arguably because the speaker declined to be more informative by referring to it as Y when the opportunity was there.

Building over these findings, the current investigation has two aims. One is to determine whether recognizable, conventional-looking objects affect participants' navigation of the *ad hoc*

[1]For full fledged linguistically-based account of this developmental effect (one that relies on scales of informativity), see Barner et al. (2011).

implicature in the same way as named unconventional objects have been shown to do in this task. To anticipate, imagine that we slightly transform the original control condition so that Y remains an unconventional, unnamed object but Z is now a recognizable, conventional-looking object. Will listeners disregard the conventional-looking object and pragmatically reason their way to choose the unnamed, unconventional-looking target referent (much like they do when there is a named unconventional object) or will they consider the conventional-looking object, which is never explicitly named, as having equal status to the unnamed, unconventional object? To put it another way, to what extent does the presence of a (readily identifiable) conventional-looking object prompt participants to exclude it from consideration upon hearing "Not the X"? Assuming that participants do eliminate a conventional-looking object from consideration in such circumstances, we then ask to what extent does this sort of information compete with a case in which the other, unconventional object is indeed named. In this way, we can capture how well (unnamed) conventional-looking objects measure up to properly named unconventional ones.

This preamble sets up what follows. Below, we present the task as it was inspired by Kronmüller et al. (2017), which was conducted in Spanish, while investigating negative reference with pre-recorded materials. We monitor listeners' moment-by-moment interpretation of negated references using a visual-world eye-tracking task. Two other critical dependent variables are participants' final referent selections and their reaction times in making selections.

MATERIALS AND METHODS

Participants

Participants include 48 native speakers of Spanish of which 26 were male. Their mean age was 22 and ages ranged from 18 to 32. Forty-three were undergraduate students from different faculties and five were professionals. They participated in exchange for 5,000 Chilean Pesos (approx. 8 USD).

Design

The experiment had four conditions that were administered within subjects. Each condition was defined by the way the test trial juxtaposed two kinds of remaining objects (what we have been calling objects Y and Z), after one named unconventional object (X) had been ruled out by the speaker (through "not the X"). We essentially turned the objects Y and Z into variables, by calling them the *second* and *third* objects. That is, the presentation of each of these two remaining objects was systematically varied, based on the following: (1) we varied whether or not the Y (the second) object, which was always unconventional looking, had been previously designated with a name, and (2) we varied whether or not the Z (the third) object, which always remained unnamed, was conventional looking. **Table 2** summarizes the experiment's design and its four condition names.

TABLE 2 | The design of the experiment's four conditions, including a description of all three objects in each upon hearing the negative reference, "not the X," which refers to the previously named unconventional object that is found in each condition.

	First object (X)	Second object (Y)	Third object (Z)
	Always unconventional and named	*Always unconventional (varies naming)*	*Always unnamed (varies conventionality)*
Condition name			
Baseline	Unconventional-named	Unconventional-unnamed	Unconventional-unnamed
Second-object-is-named (SON)	Unconventional-named	Unconventional-named	Unconventional-unnamed
Third-object-is-conventional (TOC)	Unconventional-named	Unconventional-unnamed	Conventional-unnamed
SON-vs.-TOC	Unconventional-named	Unconventional-named	Conventional-unnamed

The first condition is called "Baseline" because the negated reference of the X object leaves participants a choice between the two remaining objects (Y and Z) that are both unnamed, unconventional objects. This baseline condition is identical to the control condition in the original Kronmüller et al. (2017) paper (it equally serves as a control condition here and as a way to confirm prior results). The second-object-is-named (SON) condition refers to the case in which the negated reference of the X object leaves participants a choice between a previously named, unconventional object (Y) and an unnamed, unconventional object (Z). This condition corresponds to Kronmüller et al.'s (2017) experimental condition as described in the section "Introduction." The third-object-is-conventional (TOC) condition leaves the participant a choice between an unnamed unconventional object (Y) and a conventional-looking (Z) object (that is never explicitly named). The third condition is new to this paradigm but conceptually identical to the second-object-is-named (SON) condition. The difference between them is that we are testing whether the recognition of conventional visual information can serve as the basis of an *ad hoc* implicature. The final – SON-vs.-TOC condition – leaves the listener a choice between a named, unconventional object (Y) and unnamed, conventional-looking object (Z). This condition, which is also new to this paradigm, forces participants to choose between two "conventionalized" objects, one (Y) that was coined conversationally against an object (Z) that is assumed to be conventionalized visually. Each participant received one of four stimulus lists created by rotating each set of objects through all four conditions, such that any individual participant saw all sets. See **Figure 1** for comparable examples of a single test-trial for each condition.

Procedure

To start, the experimenter informed participants that they would be playing a game in which their task was to identify and select target pictures based on recorded directions from a previous speaker. The pictures were presented on a computer screen and selection was made by clicking on the picture with the computer's mouse. Participants were led to believe that the previous speaker was a naïve participant playing the "director" role in a communication game whose spontaneous speech was recorded while giving instructions to a "matcher" participant in a previous session. Critically, they were led to believe that during the previous session, the director and the matcher worked from different computer screens and were prevented from viewing each other's screen by a divider. During the experimental session, speech was automatically triggered by the software which also recorded the identity of the object selected and the time it took participants to make the selection.

For presentation purposes, we conceive of a single "trial" as an event in which a participant views a display and interprets the speaker's instruction to click on one of the displayed objects. A "round" is a collection of trials. **Figure 2** shows an example of a round for each condition. The trials making up a round can be subdivided into two phases: a "grounding phase" (rows A–F) and a "completion phase." The completion phase is designed to present a condition's "test trial" (one of the four in **Figure 1**) pseudo-randomly along with two other trials (those in rows G and H) such that the test trial can appear anywhere in this phase. In other words, the completion phase ultimately *includes* the "test trial" but it could arise as the 7th, 8th, or 9th trial. This was done to mask the purpose of the experiment. We performed analyses on the test trial data only, regardless of its position in the round.

The grounding phase itself consisted of four "grounding trials," used to set up the names for the test trial objects (lines A, B, C, and D), plus two "filler trials" where names were repeated or where other objects were referred to (lines E and F). In two of the grounding trials (A and B), a name is given to the object (X) that will later be negated in the test trial. In two other grounding trials (C and D), an unconventional object is either named [for the second-object-is-named (SON) or the SON-vs.-TOC conditions (see the utterance in green)], or is not named [for the baseline and the third-object-is-conventional (TOC) conditions (see the utterance in red, where an object *not appearing in the test trial* is given a name)]. Note that for the baseline and second-object-is-named (SON) conditions, the third object was also an unconventional object that would not be easy to name. In contrast, in the third-object-is-conventional (TOC) and SON-vs.-TOC conditions, the third object (Z) is a conventional-looking object (though, as always, it was never explicitly named).

Grounding trials were pseudorandomized, so that A, C, and E would always appear before B, D, and F. The rationale for this is to allow for a first mention to be more descriptive such as "the one that looks like a modern sculpture," and so that the second mention is a more concise one, such as "the sculpture." This simplification of a description is a well-known phenomenon in dialogue research (Clark and Wilkes-Gibbs, 1986).

As indicated above, the test trial was pseudo-randomly presented as one of the last three trials of a round. In each test trial, the speaker used a negative referring expression to identify the target, such as "not the sculpture" (Spanish: "no la escultura"). As can be seen in **Figure 1**, listeners saw three referential alternatives, the negated object (X), the second (Y), and third (Z) objects (for the sake of exposition, **Figure 1** presents these three at the top, the lower left, and the lower right, respectively), with each appearing with equal frequency in the grounding phase.

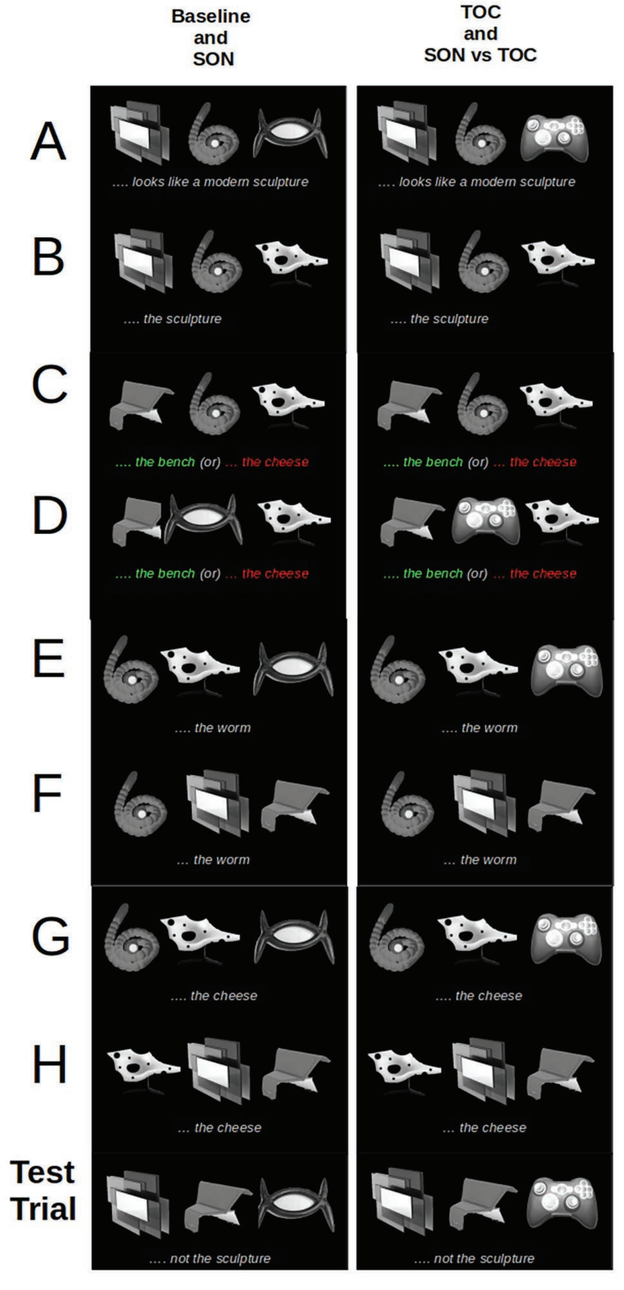


FIGURE 2 | An example of a round for each condition, including all the trials, showing the objects presented and the instructions a participant heard. The column on the left shows the objects and instructions of the baseline and second-object-is-named (SON) conditions; the column on the right depicts the objects and instructions regarding the third-object-is-conventional (TOC) and the SON-vs.-TOC conditions. In general, when the objects were named for the first time, the instructions were richer; for example, with respect to the second column in row A, the instruction would be "Click on the object that looks like a modern sculpture," whereas the second time, the instruction would be of the form "Click on the sculpture" or simply "The sculpture." Rows A and B correspond to the grounding trials that set up the negated object (X) in the test trial (the sculpture). Rows C and D depict the objects for the grounding trials among which the second object (Y) gets named or not. When it is named (in the SON and SON-vs.-TOC conditions), the instruction

(Continued)

FIGURE 2 | concerns the object circled in green (the bench). For the baseline and TOC conditions, there is another object that is named, circled in red, but it does not appear in the test trial, leaving the second object in the eventual test trial unnamed. Rows E and F are filler trials. The trials in rows G and H along with the test trial are all part of the completion phase. The trials in the completion phase appear in a pseudo-random order so that the test trial appears as either the 7th, 8th or 9th trial, so as to mask the purpose of the study.

One concern with the task is that the speaker's negated descriptions may seem uncooperative, since the description would be insufficient for distinguishing between two possible alternatives. To avoid perceptions of uncooperativeness, we added additional procedures and a cover story (similar to those used in past instantiations of this experiment in the literature). Participants were told that when the instructions were recorded, the speaker and the listener viewed their displays on different computer monitors; we also led the listener to believe that the speaker saw only two of the three alternatives that the listener saw, but that the listener did not know which two. Given this setup, listeners would have no reason to find the speaker's behavior uncooperative. In order to keep this feature salient to the participant, the experiment on occasion would request the participant to guess which was the object that the speaker was not seeing.

Materials and Apparatus

Sets of six objects were prepared for each of the 24 rounds (144 different objects in total). For each set, five of the six were unconventional objects: one for the negated object (X), one for the second (Y) object that could potentially be named, and one for the third (Z) object when it was unconventional, plus two different filler objects. The sixth was a conventional-looking object that could serve as the third (Z) object when called for. All of the images were downloaded from the Internet and were converted to grayscale so that they could not be identified by color.

We tracked listeners' eyes using an EyeLink 1,000 eyetracker (SR~Research). The system used a remote tabletop camera, allowing relatively free head movement. Gaze data were recorded at a sampling rate of ~500 Hz.

Data Analysis

We analyzed participants' proportion of selections as well as their reaction times and gaze patterns to each object. Selection data reflect the participant's final referential commitment. Eye movements, in contrast, inform us about the interpretation process in real time. Because the main interest was the relation between the second (Y) and third (Z) objects, for all inferential statistics our dependent variable is the "log-ratio" of selecting the third (Z) object over the second (Y) object across conditions: a log ratio of zero means no preference, a positive value means preference for the third (Z) object, and a negative value points to a preference for the second (Y) object.

All *p*'s in the selection and eye-movement analyses, for subjects (p1) and items (p2), were obtained using a resampling technique. We generated a permutation scheme through which a decision was made to either keep the original labeling or change the labels for all four conditions (so no data point kept its original labeling). We built 9,999 data sets based on Monte Carlo samples for all possible arrangements of the data following our permutation scheme. For the selection data, we fit a baseline-category multinomial logistic regression (Agresti, 2002) to each of these datasets and built a null hypothesis distribution with all regression coefficients against which the original coefficient was contrasted. The proportion of coefficients from the null hypothesis distribution greater than the original constitutes the *p* for a specific contrast. We take the baseline condition as the reference group in a dummy coding scheme.

Given the complexity of analyzing eye-tracking data, mainly due to the fact that the time series is categorical, and in order to avoid arbitrary identification of time windows to perform the statistical analyses, we follow a "cluster randomization" approach as it has been previously adapted to visual world eye-tracking experiments and specifically to the original version of the task we present here (see Barr et al., 2014 for a thorough explanation of this approach). In short, the algorithm identifies periods of time during which two time series diverge. Finally, reaction times were analyzed using a mixed-effects regression, with subjects and items as random factors. We include the maximal random effect structure justified by the design and that converged, which in our case was all random effects (intercepts and slopes) but excluding their correlations (Barr et al., 2013). *p*'s are obtained using a model comparison approach. All analyses and graphics were performed using R (Bates et al., 2013).

RESULTS

Selection

Figure 3 summarizes the results from the four conditions. In the baseline condition, the proportion of selection of the second (Y) object (0.49) was equal to the selection of the third (Z) object (0.49) (log-ratio = 0). In the SON condition, in contrast, the selection of the third (Z) object (0.67) was 2.1 times higher than the proportion of selections of the second (Y) object (0.32). This log ratio (0.74) is different than the zero log ratio in the baseline condition ($p1 < 0.001$, $p2 < 0.001$). In the TOC condition, the proportion of selection of the second (Y) object (0.64) was 1.9 times higher than the third (Z) object (0.34); this log ratio (−0.63) is also significantly different than the ratio in the baseline condition ($p1 < 0.001$, $p2 < 0.001$). Finally, in the SON-vs.-TOC condition, the proportion of selection of the third (Z) object (0.56) was 1.4 times higher than the second (Y) object (0.41); this log ratio (0.31) is not statistically different from the log ratio in the baseline condition ($p1 = 0.148$, $p2 = 0.115$).

Individual Differences

In order to observed individual tendencies, we present the proportion of "optimal" responses. An optimal response is defined as the speaker's likely intended object post-negative-reference (after "not the X"), based on conversational considerations.

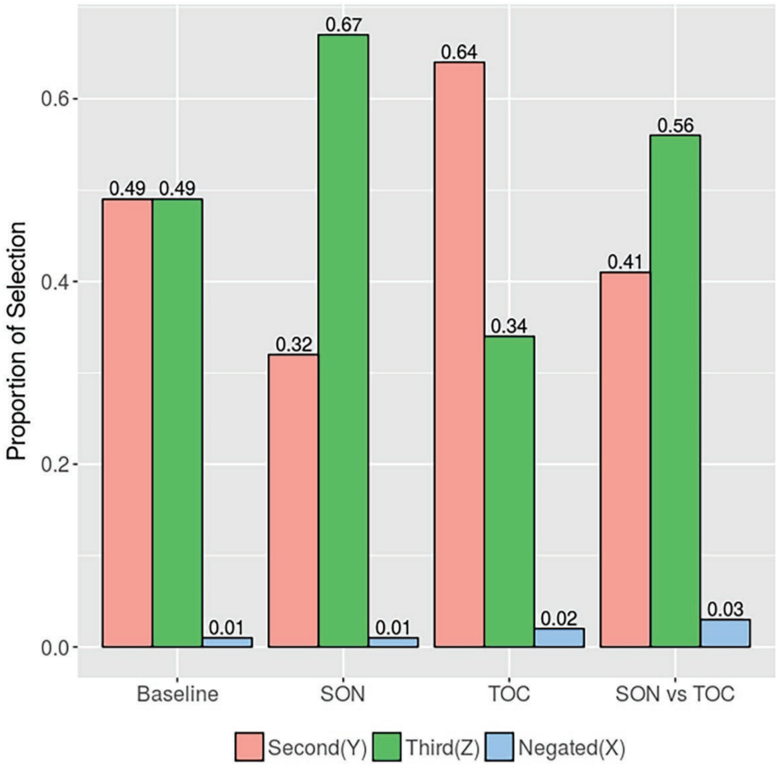


FIGURE 3 | Proportion of selection of each object per condition.

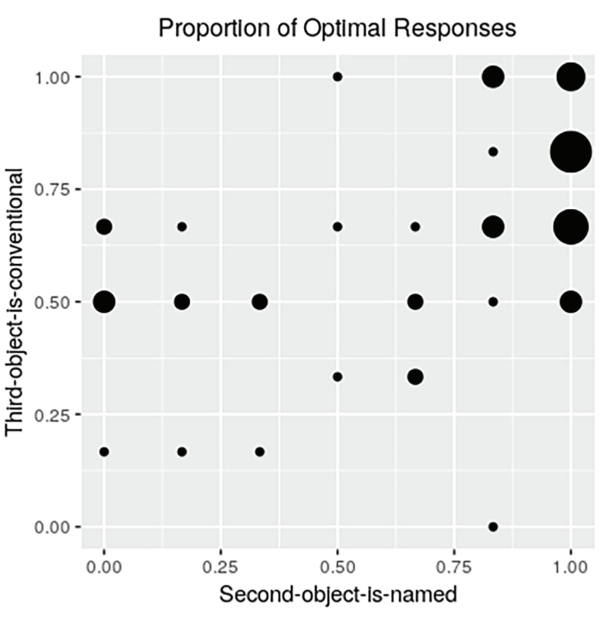


FIGURE 4 | Individual differences: proportion of optimal responses in the second-object-is-named (SON) and the third-object-is-conventional (TOC) conditions. The size of each dot on the grid represents the relative number of participants at that coordinate with respect to the proportion of optimal responses provided in the SON condition (x axis) and the proportion of optimal responses provided in the TOC condition (y axis).

In the case of the SON condition, this means selecting the third (Z) object (which is the remaining unnamed and unconventional object in this condition). In the case of the TOC condition, in contrast, the optimal alternative is the second (Y) object (which is the remaining unnamed and unconventional object, for this condition). In both of these cases, the optimal response is the one for which there is the least information. **Figure 4** presents the proportion of selections of the optimal alternative. Each dot on the grid represents the relative number of participants at that coordinate with respect to the proportion of optimal responses provided in the SON condition (x axis) and the proportion of optimal responses provided in the TOC condition (y axis). The participants in the upper right corner of the graph were consistently optimal in their responding across the two conditions. The relatively empty lower left corner of the graph represents those participants who choose both the named object in the SON condition and the conventional-looking object in the TOC condition. By visual inspection, it can be observed that there are, roughly, two groups. Most of the participants tend to make optimal responses, but there is a smaller group (on the left half of the graph) that systematically gives non-optimal responses in the SON condition, by selecting the object (Y) that had in fact been given a name, while being somewhat indifferent with respect to the TOC condition, by selecting the conventional object (Z) about half the time. The two right-most columns reveal that the SON condition provides optimal responses more consistently than the TOC condition.

Reaction Times

Here, we present the time it took – in milliseconds – to select an object upon hearing the negative reference, beginning from the offset of the noun (for example, from the offset of "sculpture" in "not the sculpture"). Reaction times show an interesting pattern. Reaction times are longer in the baseline condition (M = 3,518; SD = 1,403) compared to the SON condition (M = 2,530; SD = 805) [$\chi^2(1) = 38.523$; $p < 0.001$]. Likewise, reaction times are longer in the baseline condition when compared to those in the SON-vs.-TOC condition (M = 2,888; SD = 1,105) [$\chi^2(1) = 16.201$; $p < 0.001$]. In contrast, reaction times in the TOC condition (M = 3,302; SD = 1,305) were not statistically different from those in the baseline condition [$\chi^2(1) = 2.061$; $p = 0.151$]. Means and standard deviations were computed aggregating by subjects.

Eye Movements

Figure 5 shows the preferences for either the second (Y) or third (Z) objects from the offset of the referring expression up to 3,000 ms. As can be observed, the second-object-is-named (SON) condition reveals that there is an early preference for the third (Z) object (which is unnamed and unconventional looking) that starts at 700 ms and is sustained until the end of the time window. In the TOC condition, there is an opposing pattern, with an early preference toward the second (Y) object (which is unnamed). However, in this case, this preference does not grow monotonically until the end of the window; after their initial quick decision, participants appear to hesitate by looking back and forth between the two remaining objects. In the SON-vs.-TOC condition, there is an early preference for the (conventional looking) third (Z) object, but less than in the SON condition; much like in the TOC condition, their early preference is followed by some apparent indecision. Finally, as expected, the baseline condition is around zero throughout the entire time window.

These observations are corroborated by statistical analyses: two clusters can be identified when comparing the baseline condition against all others. For the SON condition, there is a reliable cluster that starts at 900 ms up to the end of the time window ($p1 < 0.001$, $p2 < 0.001$). For the TOC, there is a reliable cluster between 650 and 1,300 ms ($p1 = 0.036$, $p2 = 0.042$). Finally, even when there is a numerical difference between the SON-vs.-TOC and the baseline conditions, starting approximately at 1,000 ms up to 2,200 ms, there is no reliable cluster identified by the cluster randomization algorithm.

DISCUSSION

Prior work with a well-established negative reference task (Kronmüller et al.'s, 2017) has shown that, post-negative reference, participants reliably look past a previously named object in order to choose an unnamed unconventional object. This effect was replicated here through the second-object-is-named (SON) condition. We argue that participants hypothesize that, if an object can be readily referred to with a name, a speaker would have done so and that, if a speaker does not do such referring, it is taken as a clue that the negative reference is more likely referring to the unnamed unconventional-looking object. The present work extends this finding by investigating a highly similar situation in which a conventional-looking object that is never named (for example, the object that resembles a gamepad in our figures) is juxtaposed with an unnamed unconventional object. The question that we asked in this work was, would participants similarly look past the conventional object and choose

FIGURE 5 | Log ratio across time from zero milliseconds after the offset of the noun in the negated reference up to 3,000 ms. A line around zero means no preference. A positive value reflects a preference for third object (Z), and a negative value reflects a preference for the second object (Y). Each line depicts a condition.

the unnamed, nondescript alternative? The paradigm also allowed for a situation in which the named unconventional object could be set against the unnamed conventional-looking object.

The results from the third-object-is-conventional (TOC) condition indeed show that participants give optimal responses at rates that are comparable to those in the second-object-is-named (SON) condition. That said, the effect linked to the third-object-is-conventional (TOC) condition does not appear to be as a strong. This can be inferred from three results. The first is that participants' eye-tracking patterns toward the remaining unnamed unconventional object (Y) appear less resolute in the third-object-is-conventional (TOC) condition than they are to the unnamed unconventional object (Z) in the second-object-is-named (SON) condition. Participants do immediately focus on the most nondescript object in the third-object-is-conventional (TOC) condition, but they also reveal some hesitation soon afterward. This is unlike participants' reactions in the second-object-is-named (SON) condition in which they maintain their focus on the remaining unnamed unconventional object (Z) and in an increasingly monotonic fashion. The individual difference data in **Figure 4** provide further supporting evidence showing that the SON condition provides more resolute decision making than the TOC condition. There, one sees 100% or near 100% optimal choice making for a majority of the participants (the two rightmost columns of **Figure 4**), which reflects optimal performance on the second-object-is-named (SON) condition, while optimal performance for the third-object-is-conventional (TOC) condition is less common for the top two rows of **Figure 4**.

The second result is that reaction times in the third-object-is-conventional (TOC) condition were as slow as in the baseline conditions. In contrast, in the second-object-is-named (SON) condition, reaction times were relatively fast. And, finally, the third result is that when the two sorts of cases are forced to compete in the SON-vs.-TOC condition, one can detect that there is a slight (though not statistically reliable) tendency to favor the unnamed conventional object (Z) as the participant's choice selection. Both in terms of selections and in terms of eye-tracking patterns, listeners tend to look past the named object and to choose the unnamed conventional-looking object as the speaker's referent. This summary, of course, refers to overall preferences for this particular task. Clearly, there are many participants who do not use the speaker's cues to make what we refer to as the optimal response. It is also important to keep in mind that the conventional-looking object is never given a mutually manifest name. It will be useful for future work to determine the extent to which a *named*, conventional-looking object determines performance on this task. The aim of that work will be to determine whether the effects of naming and conventional appearance are additive.

The regularity of these results is quite remarkable once one considers that participants are simply receiving a negative sentence in the form of "not the X," which leaves two options and much else to be determined. To come up with what this paradigm considers to be the optimal response, listeners are arguably reasoning that the speaker could have made a more direct and informative, i.e., more facilitative, utterance [e.g., *(Click on) the Y object*] but did not. In light of this, the addressee is justified in assuming that the object in such an unspoken – and potentially more informative utterance – is *not* the one that the speaker had intended to point out. This leads the listener to conclude that the speaker did not intend to refer the *Y* object in the second-object-is-named (SON) condition; likewise, it leads a listener to the conclusion that the speaker did not intend to refer to the *Z* object in the third-object-is-conventional (TOC) condition.

Ad hoc Implicature

The current findings show the extent to which pragmatic reasoning need not rely on linguistic features. As we noted in the section "Introduction," we argue that optimal choices result from a particularized or *ad hoc* implicature in which listeners consider the contents of a two member category created by the negative reference; more specifically, participants look past the object that has the potential to be readily informative because the speaker did not mention it. These findings edify our understanding of *ad hoc* implicature-making in three important ways.

The first is that a listener's *ad hoc* pragmatic reasoning here leads to an optimal reference at above-chance levels in the SON and TOC conditions even though the speaker is assumed to be viewing just two of the three objects in front of the listener. To make a non-random reference, the listener actually needs to *infer* which objects appear on the speaker's screen, based on what was said (or observed). In other words, the listener needs to generate a speaker's epistemic state in order to justify the optimal choice. This is more complicated than what occurs in classic *ad hoc* implicature tasks, in which a speaker refers to a category of, say, three similar drawings and the speaker is assumed to have the same, stable view as the listener. Nevertheless, the relatively reliable results reported here indicate that listeners use a procedure that is similar to those found with other *ad hoc* implicatures, in which participants consider what could have been said but was not.

The second is that optimal responding need not be determined uniquely by prior actions taken by the speaker. The mere presence of a conventional looking object, one that occasionally and namelessly appears across a round (as is the case for the conventional looking object in the TOC condition), also encourages listeners to assume that the negative reference leads to an optimal response, i.e., to choose the most non-descript object. This shows how *ad hoc* implicature-making is opportunistic; a listener will seek out any sort of evidence in an effort to identify an alternative that can make distinctions with regard to informativeness. When the salient conventional-looking object is not referred to in the third-object-is-conventional (TOC) condition, it is presumably a clue to the participant that it is not on the speaker's screen.

Finally, this is the only paradigm we are aware of in which *ad hoc* procedures operate reliably in a wholly negative space. Participants begin their calculation through a negative reference and then disprefer one object out of the remaining two based on their interlocutors' conversational history (in the SON condition) or on the salience of a potentially nameable object (in the TOC condition). Optimal performance is not based on contrasts between highly similar objects (such as smiley faces) or on the categories that the objects can spontaneously belong to. Overall, participants' ability to find optimal responses in this difficult context is impressive.

Conclusions

The investigation here is exemplary of the kind of work that is needed to better understand the role played by conventionalizations in language as they are employed in utterance understanding. There is much else left to do. Other questions that experiments in this genre can answer are the following: are there indeed isolable procedures linked to conventional-looking objects? One can also ask how does performance on this task develop? These and other experimental pragmatic questions can be addressed by turning one's attention to conventionalized meanings in dialogue and to conventional-looking objects.

In conclusion, a negated reference in the current paradigm forces a listener to rely almost exclusively on contextual information in order to infer communicative intentions. We show how the negative reference *Not the X* triggers the creation of an *ad hoc* category and a pragmatic process through which the listener needs to evaluate two alternatives, with the optimal response amounting to recognizing which object could have been identified in an informative way but was not. Ultimately, listeners needed to identify the least mutually-recognizable object of two in a task that started with a negative expression. It appears that unconventional-looking objects having temporarily shared names carry slightly more weight than conventional-looking objects that are never explicitly referred to.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of ethical guidelines of the Comité Asesor de Bioética del Fondo Nacional de Desarrollo Científico y Tecnológico de Chile. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Comité Asesor de Bioética del Fondo Nacional de Desarrollo Científico y Tecnológico de Chile.

AUTHOR CONTRIBUTIONS

This represents the authors' third joint project. EK and IN conceived the current study. EK designed it, ran it and analyzed the results. The joy in appreciating the findings and writing up the paper was equally distributed.

FUNDING

Financial support for this work came from *Fondo Nacional de Desarrollo Científico y Tecnológico – CONICYT*, Chile to EK and from *l'Institut des Sciences Cognitives* – Marc Jeannerod (UMR 5304), France to IN.

REFERENCES

Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: Wiley.

Barner, D., Brooks, N., and Bale, A. (2011). Accessing the unsaid: the role of scalar alternatives in children's pragmatic inference. *Cognition* 118, 84–93. doi: 10.1016/j.cognition.2010.10.010

Barr, D. J., Jackson, L., and Phillips, I. (2014). Using a voice to put a name to a face: the psycholinguistics of proper name comprehension. *J. Exp. Psychol. Gen.* 143, 404–413. doi: 10.1037/a0031813

Barr, D. J., and Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *J. Mem. Lang.* 46, 391–418. doi: 10.1006/jmla.2001.2815

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Barsalou, L. W. (1983). *Ad hoc* categories. *Mem. Cogn.* 11, 211–227. doi: 10.3758/BF03196968

Bates, D., Maechler, M., and Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes. Retrieved from http://CRAN.R-project.org/package=lme4 (R package version 0.999999-2).

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *J. Mem. Lang.* 61, 171–190. doi: 10.1016/j.jml.2009.04.003

Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi: 10.1016/0010-0277(86)90010-7

Grice, P. (1989). *Study in the way of words*. Cambridge, MA: Harvard university Press.

Katsos, N., Cummins, C., Ezeizabarrena, M. J., Gavarró, A., Kraljević, J. K., Hrzica, G., et al. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proc. Natl. Acad. Sci. USA* 113, 9244–9249. doi: 10.1073/pnas.1601341113

Kronmüller, E., and Barr, D. J. (2007). Perspective-free pragmatics: broken precedents and the recovery-from-preemption hypothesis. *J. Mem. Lang.* 56, 436–455. doi: 10.1016/j.jml.2006.05.002

Kronmüller, E., and Barr, D. J. (2015). Referential precedents in spoken language comprehension: a review and meta-analysis. *J. Mem. Lang.* 83, 1–19. doi: 10.1016/j.jml.2015.03.008

Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., and Barr, D. (2017). The positive side of a negative reference: the delay between linguistic processing and common ground. *R. Soc. Open Sci.* 4:160827. doi: 10.1098/rsos.160827

Metzing, C., and Brennan, S. E. (2003). When conceptual pacts are broken: partner-specific effects on the comprehension of referring expressions. *J. Mem. Lang.* 49, 201–213. doi: 10.1016/S0749-596X(03)00028-7

Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* 78, 165–188. doi: 10.1016/S0010-0277(00)00114-1

Noveck, I. A. (2018). *Experimental pragmatics: the making of a cognitive science*. Cambridge: Cambridge University Press.

Papafragou, A., and Tantalou, N. (2004). Children's computation of implicatures. *Lang. Acquis.* 12, 71–82. doi: 10.1207/s15327817la1201_3

Stiller, A. J., Goodman, N. D., and Frank, M. C. (2015). *Ad-hoc* implicature in preschool children. *Lang. Learn. Dev.* 11, 176–190. doi: 10.1080/15475441.2014.927328

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 August 2019
doi: 10.3389/fpsyg.2019.01782

Check for updates

Brain Inhibitory Mechanisms Are Involved in the Processing of Sentential Negation, Regardless of Its Content. Evidence From EEG Theta and Beta Rhythms

*David Beltrán[1,2], Yurena Morera[1,2], Enrique García-Marco[1,3,4] and Manuel de Vega[1,2]**

[1] Instituto Universitario de Neurociencia, Universidad de La Laguna, San Cristóbal de La Laguna, Spain, [2] Departamento de Psicología Cognitiva, Universidad de La Laguna, San Cristóbal de La Laguna, Spain, [3] Centro Asociado de La Laguna, Universidad Nacional de Educación a Distancia, Madrid, Spain, [4] Facultad de Psicología, Universidad Europea de Canarias, Santa Cruz de Tenerife, Spain

OPEN ACCESS

Edited by:
*Katharina Spalek,
Humboldt University of Berlin, Germany*

Reviewed by:
*Jana Lüdtke,
Freie Universität Berlin, Germany
Francesca Delogu,
Saarland University, Germany*

***Correspondence:**
*Manuel de Vega
mdevega@ull.edu.es*

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *21 January 2019*
Accepted: *17 July 2019*
Published: *08 August 2019*

Citation:
Beltrán D, Morera Y, García-Marco E and de Vega M (2019) Brain Inhibitory Mechanisms Are Involved in the Processing of Sentential Negation, Regardless of Its Content. Evidence From EEG Theta and Beta Rhythms. Front. Psychol. 10:1782. doi: 10.3389/fpsyg.2019.01782

The two-step process account of negation understanding posits an initial representation of the negated events, followed by a representation of the actual state of events. On the other hand, behavioral and neurophysiological studies provided evidence that linguistic negation suppresses or reduces the activation of the negated events, contributing to shift attention to the actual state of events. However, the specific mechanism of this suppression is poorly known. Recently, based on the brain organization principle of neural reuse (Anderson, 2010), it has been proposed that understanding linguistic negation partially relies upon the neurophysiological mechanisms of response inhibition. Specifically, it was reported that negated action-related sentences modulate EEG signatures of response inhibition (de Vega et al., 2016; Beltrán et al., 2018). In the current EEG study, we ponder whether the reusing of response inhibition processes by negation is constrained to action-related contents or consists of a more general-purpose mechanism. To this end, we employed the same dual-task paradigm as in our prior study—a Go/NoGo task embedded into a sentence comprehension task—but this time including both action and non-action sentences. The results confirmed that the increase of theta power elicited by NoGo trials was modulated by negative sentences, compared to their affirmative counterparts, and this polarity effect was statistically similar for both action- and non-action-related sentences. Thus, a general-purpose inhibitory control mechanism, rather than one specific for action language, is likely operating in the comprehension of sentential negation to produce the transition between alternative representations.

Keywords: sentential negation, two-step account, response inhibition, theta rhythms, beta rhythms, inhibition reuse

INTRODUCTION

Negation—as instantiated by operators like *not* and *no*—belongs to the special class of linguistic devices whose understanding in sentential contexts implies representing at least two different, often opposed alternatives. According to the so-called two-step process of negation (e.g., Kaup and Zwaan, 2003), negative sentences (e.g., *Today is not a bright day*) are semantically more complex than the corresponding affirmative sentences (e.g., *Today is a bright day*). The explanation is that the latter expresses only one idea, which corresponds to the actual state of affairs, whereas a negative sentence induces the reader/listener to represent the negated situation (e.g., *a bright day*) as well as the actual one (e.g., *a cloudy or dark day*). This conception is clearly supported by a recurrent finding reported in the literature: the comprehension of negative statements generally demands more cognitive resources and processing time than the comprehension of affirmative sentences (for reviews, Wason and Johnson-Laird, 1972; Kaup, 2001; Tian and Breheny, 2016; Papeo and de Vega, 2019).

An important aspect in the literature of the two-step account is that the representation of the negated events is temporary, since it is rapidly suppressed and replaced by the representation of the actual events. No doubt, managing two representations in negations (e.g., suppressing one and activating and keeping the other) requires efficient processes that often have been neglected. The present study tries to examine one of these processes, proposing that the response inhibition system could be responsible to make the transition between the initial and the actual representation derived from negative polarity sentences, by suppressing the former. Moreover, this paper posits that response inhibition is recruited for processing sentential negation, regardless of its content.

From Representations to Processes

Cognitive research on negation has traditionally focused on the temporal dynamic of the two underlying representations, following the prevailing two-step process account. For instance, for the aforementioned negative sentence "Today is not a bright day," this model proposes that a representation of the denied situation is activated first as if the negative operator had been removed, and hence creating a similar meaning representation as the affirmative counterpart (e.g., *Today is a bright day*). Next, in a second step, the negative operator starts to be integrated into the sentence meaning, resulting in deactivation of the initial representation, and a later replacement by the representation of the actual state of affairs (e.g., *A cloudy or dark day*). There are competing models to explain negation processing, and also empirical findings that question some of the assumptions of the two-step account, specially the one stipulating that the first step is mandatory (e.g., Mayo et al., 2004; Giora et al., 2007; Khemlani et al., 2012; Tian and Breheny, 2016). At least in some cases, sentential negation seems to be processed in the same way as affirmative sentences. For instance, world-knowledge violations in negative (e.g., *Zebras are not stripy*) and affirmative form (e.g., *Ladybirds are stripy*) induce the same N400 modulations and do not show any evidence of an additional processing step (Dudschig et al., 2018). Also, some studies described in the next section reported that negation induces a disembodiment effect in action language very early, as measured by grip force (Aravena et al., 2012) or corticospinal excitability measures (Papeo et al., 2016), suggesting a single-step processing of sentential negation.

In any case, all the theoretical and empirical approaches share a concern on what is represented and when, and also on how pragmatic factors—background knowledge and context information—modulate the whole process of sentence meaning comprehension (e.g., Beltrán et al., 2008; Nieuwland and Kuperberg, 2008; Dale and Duran, 2011; Orenes et al., 2014, 2016). However, negative and affirmative sentences differ not only in the number of alternatives (or representations) they invoke but also in the operations (or processes) recruited to manage these representations. We think that the analysis of these processes has been somehow neglected by previous studies, with a few recent exceptions (de Vega et al., 2016; Beltrán et al., 2018; Dudschig and Kaup, 2018).

Let us focus on the activation–inhibition processes proposed by the two-step account, which are inferred from the results obtained with experimental paradigms such as the probe recognition task (e.g., MacDonald and Just, 1989; Kaup, 2001; Kaup and Zwaan, 2003; Kaup et al., 2006). In this task, a sentence (or a short paragraph) is followed by a *probe* (a word or a picture), and participants have to recognize whether this *probe* was previously mentioned; in other versions of the task, participants simply name the probe aloud. The latency to the *probe* is the key measure, which is taken as an index of activation for the corresponding concept. A common pattern obtained was as follows: when the *probe* was shown shortly after the sentence, the time to recognize or name it was the same regardless of the polarity of the sentence (e.g., *The door is [not] open*), whereas when the interval between the sentence and the *probe* was large, then the recognition (or naming) latencies were larger for negative than for affirmative sentences (e.g., Kaup et al., 2006). The latter result is usually interpreted as reflecting the suppression or inhibition of the negated concept and hence as a demonstration of the second step for negation processing. Accordingly, one key feature of negation, relative to affirmative sentences, is that, over time, it recruits additional processes. But what are these additional processes? Based on the empirical consequences of negation, most researchers agree that the function of these processes is twofold: inhibiting (or suppressing) the negated content and updating (or activating) a representation of the actual situation. Still, they have not usually gone beyond this general description. Which is the brain machinery underlying these processes? Is it a neural network specifically involved in the syntactic processing of sentential negation? Or, by contrast, is it a general-purpose inhibitory control network, primarily involved in monitoring alternative actions and reused to managing alternative linguistic meanings in sentential negation?

As we will see in the next section, a first approach to the neural bases of understanding sentential negation derives from the embodiment research program, aimed to specify how sensory-motor systems contribute to represent sentence meaning. Specifically, several researchers applied the embodied approach to contrast affirmative and negative action sentences.

Embodied Research on Negation

The embodied approach to language comprehension posits that meaning is grounded in the activity of non-linguistic systems (for recent reviews, Barsalou, 2016; García and Ibáñez, 2016). Crucially, there is extensive empirical evidence of embodied effects during the comprehension of action-related language, demonstrating that it partially relies on the activation of the motor mechanisms. For instance, the behavioral paradigm action-sentence compatibility effect (ACE) has shown that understanding action sentences interacts (facilitating or interfering) with performance in a concurrent matching motor task (e.g., Glenberg and Kaschak, 2002). Also, motor and premotor cortex processes associated with action language have been revealed by neuroimaging (e.g., Tettamanti et al., 2008; Moody and Gennari, 2010; Tomasino et al., 2010; de Vega et al., 2014), electroencephalography (e.g., Aravena et al., 2010; van Elk et al., 2010; Moreno et al., 2013, 2015), non-invasive brain stimulation (e.g., Buccino et al., 2005; Tomasino et al., 2008; Papeo et al., 2009), and brain-injured patient studies (e.g., Boulenger et al., 2008; Herrera et al., 2012).

The most remarkable phenomenon for the purpose of this article is that the presence of a negative operator in action-related statements produces a "disembodiment" effect by reducing motor activation, compared to their affirmative counterparts. Thus, behavioral studies have demonstrated that negation reduces peripheral motor activity underlying the semantics of action language (Aravena et al., 2012; Bartoli et al., 2013; Foroni and Semin, 2013). For instance, in Aravena et al.'s (2012) study, the participants kept in their right hand a grip force sensor while listening to affirmative or negative action sentences (*At the gym, Fiona lifts [doesn't lift] the dumbbells*). The results showed that the grip force does not differ between affirmative and negative sentences during the first 200 ms after listening to the action verb. However, from this moment on, the grip force steadily increased until the end for affirmative action sentences, whereas it does not differ from baseline for negative action sentences. A recent study also reported that reading negated sentences referred to manual actions (e.g., *you don't sign it*) interferes with typing the verb, whereas reading other negative statements referred to non-manual action (*you don't talk to her*) or non-motor events (*you don't believe it*) does not interfere with typing (García-Marco et al., 2019).

The effect of negation on manual action language has also been reported in some studies using neural measures. Thus, single-pulse TMS applied over the hand motor cortex revealed modulations in corticospinal excitability when reading affirmative manual verbs (*I write*), but not when reading negated manual verbs (*I don't write*); by contrast, abstract verbs did not modulate motor excitability regardless of their polarity (*I wonder/I don't wonder*) (Liuzza et al., 2011; Papeo et al., 2016; Experiment 1). Also, neuroimaging studies have shown increased activation of the motor and premotor cortex during the comprehension of affirmative action sentences and considerable reduction of these activations while understanding their negative counterparts (Tettamanti et al., 2008; Tomasino et al., 2010). Thus, overall, these investigations suggest that the motor system is recruited to process the meaning of affirmative action sentences, whereas it is deactivated or inhibited during the processing of negative action sentences.

Most of the above studies mainly reported disembodiment effects of negation in action-related linguistic contents, although specific effects of negation in non-action domains have also been obtained by Tettamanti et al. (2008), who reported a deactivation of the posterior cingulate cortex in negative abstract sentences compared to their affirmative counterparts. However, beyond the general or the content-specific neural deactivations induced by negation, none of the above studies proposed a general brain mechanism that could be responsible for these negation-induced deactivations. An interesting exception was a recent neuroimaging study, using a pattern analysis algorithm to reveal that affirmative and negative sentences regardless of their specific content differentially modulate the activation of several brain areas, including the left dorsolateral, the medial frontal cortex, the anterior and middle cingulate gyrus, and the precuneus (Ghio et al., 2018). However, in the same study the authors also found that negative sentences uniquely modulate content-specific brain areas for concrete sentences (left posterior temporal gyrus, left angular gyrus, right inferior frontal gyrus (IFG), and right superior frontal gyrus) and for abstract sentences (left temporal pole, right medial temporal lobe, right precuneus, and cerebellum), indicating that the impact of negation might be highly distributed and content-dependent.

Neural Inhibition in Negated Action Sentences

Recently, it has been proposed that one of the neural mechanisms underlying the processing of negation is the response inhibition network of the brain (de Vega et al., 2016; Papeo et al., 2016; Beltrán et al., 2018). The inhibition system is a well-known network that includes prefrontal structures, such as the right IFG, and the pre-supplementary motor area (pre-SMA) among others, which are typically involved in inhibition and control processes observed in several experimental paradigms such as the Go–NoGo or the Stop signal (Aron et al., 2014, for a review). When the EEG is recorded during the performance of these tasks, response inhibition produces robust signatures. Thus, refraining from responding in NoGo trials, in the context of a prepotent response requested in the frequent Go trials, is associated with increased power in fronto-central theta band (4–7 Hz) rhythms (Nigbur et al., 2011; Huster et al., 2013; Harper et al., 2014), as well as enhanced amplitude of the N1, N2, or P3 components of the ERPs (Bokura et al., 2001; Maguire et al., 2010).

To explore how these inhibition signatures are modulated by sentential negation, de Vega et al. (2016) asked participants to read hand-action sentences with affirmative or negative polarity (i.e., *Now you will [will not] cut the bread*), with an embedded Go–NoGo task. As expected, the analysis of the EEG signal provided a strong increase of power in fronto-central theta rhythms in NoGo trials, compared to Go trials, indexing motor inhibition in the former. Crucially, this effect was qualified by the sentence polarity, given the fact that negative sentences diminished NoGo theta rhythms compared to affirmative sentences, whereas no effect of polarity was observed on Go trials. This Go/NoGo × polarity interaction suggests that response suppression and linguistic negation may share inhibitory mechanisms. In another EEG study, Beltrán et al. (2018) asked participants to read the same affirmative and negative action sentences, while performing a Stop Signal task (SST). In the typical SST procedure, participants receive a Go cue in every trial, but in some trials, after a variable delay, they also receive a Stop signal, indicating prompting the suppression of the underway response. The Stop–Signal delay (SSD) contingently varies from trial to trial so as to produce around 50% successful stops. An interaction was obtained between sentence polarity and performance in stop trials (success vs. failure) in the N1 component, an early signature of inhibition processes, consisting of larger amplitude for successful trials with negative sentences than for successful trials with affirmative sentences, whereas no polarity effect was found in unsuccessful trials. The source of these modulatory effects of polarity was the right IFG, a prominent region in the neural network of response inhibition (Aron et al., 2014). Convergently, the estimated stop-signal reaction time showed that participants were significantly faster at inhibiting responses in the context of affirmative sentences than in the context of negative sentences.

Finally, in the aforementioned study by Papeo et al. (2016, Experiment 2), the authors measured the motor silent period, a marker of activity in the GABAergic system, following stimulation of the motor cortex while contracting the right-hand muscles. They obtained larger silent period while processing negated action sentences, compared to their affirmative counterparts, concluding that negation not only reduces motor activity but also recruits inhibitory processes.

A complementary hypothesis of negation has been proposed by Dudschig and Kaup (2018), according to which negation would rely on conflict-monitoring processes to cope with the two alternative representations. In their experiment, they used an analog of the Simon task in which the participants had to press the requested right or left key in the keyboard, when reading affirmative ("now right" or "now left") or negative ("not right" or "not left") prompts. They recorded the ERP lateralized readiness motor potential (LRP) that allowed exploring the time course of the initial (counterfactual) representation of negative statements (e.g., "*left*" in "*not left*") and the final (factual) representation (e.g., "*right*" in "*no left*"). Initially, the LRP corresponded to the counterfactual meaning (e.g., right hemisphere activation in "*not left*"), and later on it reversed indexing the factual representation (e.g., left hemisphere activation in "*not left*"). These results with a very specific type of linguistic negation clearly support the two-step process theory and, according to the authors, also indicate a conflict monitoring process similar to that reported in the studies with the Simon task.

The results of the above experiments go beyond the "disembodiment" effects of negation previously reported, demonstrating for the first time that linguistic negation consumes neural resources of response inhibition and/or conflict monitoring. However, the experiments just employed hand-action sentences and it is not clear whether inhibition is a general feature of negation or it is only recruited when negation is applied to action contents.

The Current Research

This study deals with an important question that remains unanswered. Based on the reported evidence, let us assume the two-step account as a default hypothesis for the processing of sentential negation. Let us also accept that the neural mechanism of inhibitory control is involved to some extent in the processing of some negative sentences. Given these premises, does inhibition work locally just on the motor system and, consequently, does it exclusively support the processing of negated action language? Or, alternatively, is inhibitory control a general-purpose mechanism operating in the processing of all negative sentences? The only evidence of the latter was the aforementioned Ghio et al. (2018) study that reported that negation, independently of its semantic content, modulates a broad neural network, with "syntactic and cognitive control" functions. In this study, we go one step further to test the generality of control inhibitory processes in sentential negation, recording neurodynamical rather than neuroanatomical data, which provide specific signatures of inhibition with fine-grained temporal resolution. To this aim, we performed an EEG study with the same dual-task paradigm as the one employed by de Vega et al. (2016). Namely, participants read affirmative and negative sentences for comprehension, while simultaneously performing a Go–NoGo task, receiving the corresponding cue 300 ms after the verb onset and 800 ms after the appearance of the polarity marker (see **Figure 1**). Note that this timing implies that the impact of negation on response inhibition was examined quite early, in a stage in which the negative marker and the verb are being integrated and before completing the processing of the whole sentence. This could be a critical moment to register neural markers of the inhibitory control processes, which could be responsible either for suppressing the initial representation of the negated situation—in the two-step account—or for preventing its activation—in an incremental single-step view.

Time–frequency analysis was time-locked to the Go–NoGo cue, with a focus on modulations in fronto-central theta rhythms (4–7 Hz). We also analyzed modulations in the right-frontal beta rhythms (13–30 Hz), which can also be sensitive to inhibitory control processes according to some recent studies (Zhang et al., 2008; Klepp et al., 2015; Wagner et al., 2018). We expect that negative sentences, compared to affirmative ones, modulate these rhythms especially when they appear in the context of response inhibition (NoGo trials). The rationale of this prediction, already probed in de Vega et al.'s (2016) study, is that the ongoing processing of negative sentences interacts with the response suppression triggered by the NoGo cue, given the fact that both share neural resources of inhibitory control. Then, this interaction is expected to happen during the developing of the first processing step of negation, namely, during the activation of the mental representation for negated information.

Critically, and unlike in previous studies, we manipulated the linguistic content including both motor action sentences (e.g., *Now you will [will not] cut the bread*) and mental events sentences (e.g., *Now you will [will not] wish a surprise*). We expect to find increase in power of theta and beta rhythms for NoGo trials, but these inhibitory markers will be also modulated by negation. What is more important, we will be able to answer our main question. If the modulatory effects of negation on these neural signatures occur just for action language, then we will have a local content-specific recruitment of the inhibitory control system. Namely, the response inhibition network would only operate on the motor cortex and therefore would only modulate negative action language. This possibility exists, given the fact

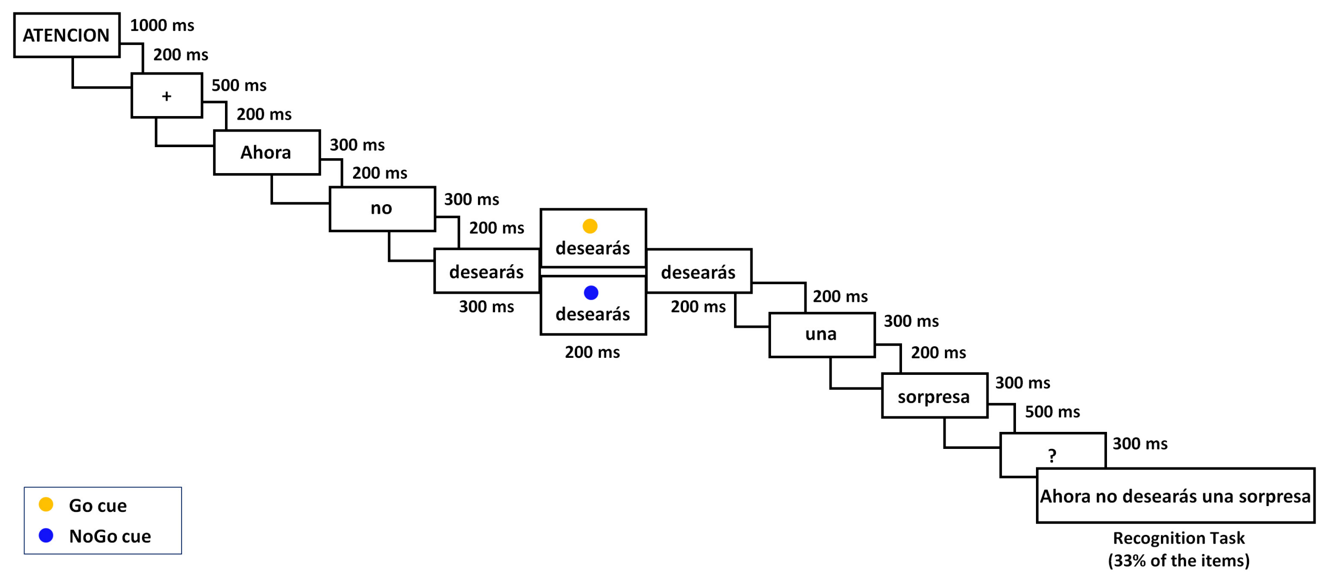


FIGURE 1 | Outline of an experimental trial with a negative mental sentence (ATTENTION/Now/you will not/wish/any/surprise/?/Now you will not wish any surprise); 70% of trials received a Go cue (yellow circle) and 30% received a NoGo cue (blue circle).

that many studies on sentential negation, registering motor performance or corticospinal excitability, found disembodiment effects for action language and null effects for non-action language (Aravena et al., 2012; Bartoli et al., 2013; Foroni and Semin, 2013). By contrast, if the modulatory effects of negation are shared by motor and mental contents, then we may support the hypothesis that inhibitory control processes are a general mechanism underlying sentential negation. Note, however, that even if theta rhythms were equally modulated by action and non-action negative sentences, this would not preclude the possibility that other content-specific networks are differentially affected by the negation marker, even though our EEG time–frequency analysis cannot dissociate them.

MATERIALS AND METHODS

Participants

A total of 27 undergraduate students of psychology participated in this experiment (19 females; age range, 19–26 years old). All participants gave written informed consent and received course credit for their participation. All were neurologically healthy, right-handed native Spanish speakers and had normal or corrected-to-normal eyesight. The study was approved by the Ethics Committee of the University (Register CEIBA2014-0126, Comité de Ética de la Investigación y Bienestar Animal. Vicerrectorado de Investigación y Transferencia de Conocimiento. Universidad de La Laguna. 38071, La Laguna, Santa Cruz de Tenerife, Spain).

Materials

A total of 532 five-word experimental sentences were created; 266 with motor action verbs (involving the use of hands) and 266 with mental verbs (involving cognitive or mental processes). The motor and the mental verbs were matched in frequency and length (see **Supplementary Table 1**), according to the EsPal database (Duchon et al., 2013), whereas, as expected, they differed in imageability, $t(28) = 11.786$, $p < 0.001$. Each verb appeared in 12 or 13 different sentences across the whole set of stimuli. Eighty additional filler sentences were also created, differing from the experimental ones in using different temporal adverbs and types of verbs. For each experimental sentence, there were two polarity versions: affirmative and negative. About one third of the sentences were followed by a recognition task to encourage participants to pay full attention to their meaning. This task consisted of a literal repetition of the previous sentence (response *yes*) or an altered version in which the polarity marker, the verb or the noun, differed from the original version (response *no*). **Table 1** shows examples of materials.

Design and Procedure

A repeated measure experimental design with 2 Cue (*Go*/*NoGo*) × 2 Polarity (*affirmative*/*negative*) × 2 Content (*motor*/*mental*) was employed. Each trial consisted of a sentence presented on a 24-inch monitor one word at a time, followed each by a blank screen; also, at a given moment, a Go or NoGo cue appeared over the sentence verb as **Figure 1** illustrates. All events in a trial were controlled by means of E-prime software (version 2.1; Psychology Software Tools). Note that 300 ms after the verb onset, the Go/NoGo cue appeared above the word as a yellow or a blue circle, respectively, during 200 ms and the verb remained for an additional 200 ms (namely, a total of 700 ms). In *Go* trials (70%), in response to the yellow circle cue, participants should press with their right-hand index finger the letter "l" on the keyboard, which was covered with a yellow sticker. In *NoGo* trials (30%), cued by the blue circle, participants should refrain from pressing any key. One third of the trials were followed by a verification sentence that was presented 800 ms after the sentence last word. The verification task consisted of responding whether or not the sentence matched the previous one by pressing with left-hand middle or index finger one of two keys labeled as "yes" or "no," respectively (corresponding to the 1 and 2 numbers in the upper left part of the keyboard). The verification sentences were correct in 50% of trials.

The structure of the session was as follows. First, the participants received instructions of the experiment followed by 16 practice trials. Thereafter, they were given six blocks of Go/NoGo trials. Four of these blocks included 101 trials each: 44 with affirmative, 44 with negative, and 13 with filler sentences; the other two blocks included 104 trials each; 45 with affirmative, with 45 negative, and 14 with filler sentences. The polarity of sentences was counterbalanced among participants, namely, a given content was presented as affirmative for half of the participants and as negative for the rest. Half of the participants began the experiment with a set of three blocks containing motor sentences (and fillers) followed by another set of three blocks including mental sentences (and fillers), and for the remaining participants, the order of the sets was reversed. Within each set, the blocks were randomly ordered for each participant, and within each block, the trial order was also randomized. The ratio of Go/NoGo trials (70%/30%) remained constant in all blocks of the experiment. The assignment of sentences to Go and NoGo trials was fixed (not counterbalanced), although the main lexical variables were matched for verbs (frequency and length) and nouns (frequency, length and imageability) in both kinds of trials, as **Table 2** shows. The duration of the experiment was 1 h approximately. Correct response reaction times and accuracy were collected for both the *Go/NoGo* task and the verification task.

TABLE 1 | Examples of experimental and filler sentences (with literal translations into English in parentheses).

Motor action:

Ahora sí [no] cortarás el pan (Now you will [will not] cut the bread)

Possible control questions*

Ahora sí cortarás el pan (Now you will cut the bread)

Ahora no cortarás el pan (Now you will not cut the bread)

Ahora sí comprarás el pan (Now you will buy the bread)

Ahora sí cortarás el queso (Now you will cut the cheese)

Mental action:

Ahora sí [no] desearás una sorpresa (Now you will [will not] wish a surprise)

Possible control questions*

Ahora sí desearás una sorpresa (Now you will wish a surprise)

Ahora no desearás una sorpresa (Now you will not wish a surprise)

Ahora sí prepararás una sorpresa (Now you will prepare a surprise)

Ahora sí desearás un consejo (Now you will wish an advice)

**A control question followed the experimental sentences in one of four versions: correct for affirmative sentences, correct for negative sentences, incorrect verb, or incorrect noun.*

TABLE 2 | Mean scores of lexical frequency, length (number of letters), and imageability of the verbs and the noun used in Go and NoGo trials.

	Verb		Noun	
	Go	**NoGo**	**Go**	**NoGo**
Motor sentences				
Frequency	0.93	0.93	0.85	0.91
Length	6.28	6.28	6.47	6.23
Imageability	5.37	5.37	6.03	6.09
Mental sentences				
Frequency	0.99	0.99	1.60	1.68
Length	6.57	6.57	7.33	7.37
Imageability	3.26	3.26	4.44	4.49

EEG Recording and Pre-processing

EEG and EOG signals were recorded using Ag/AgCl electrodes mounted in elastic Quick-caps (Compumedics). EOG signal was measured from two bipolar channels: one from two electrodes placed at the outer canthus of each eye and the other from two electrodes above and below the left eye. EEG signal was recorded from 60 electrodes arranged according to the standard 10–20 system, with additional electrodes placed at cb1/cb2 and also on the left and right mastoids (M1/M2). All EEG electrodes were referenced online to an electrode at vertex and re-referenced offline to an average reference. EEG and EOG signals were amplified at 500 Hz sampling rate using Synamp2 amplifier (Neuroscan; Compumedics), with high- and low-pass filters set at 0.05 and 100 Hz, respectively. EEG electrode impedance was kept at <5 kΩ. EEG data preprocessing and analysis were conducted using Fieldtrip Toolbox (Oostenveld et al., 2011). Trial epochs were extracted from 2.5 s precue (Go/NoGo signal) onset to 2.5 s post cue onset, resulting in 5-s epochs. Trials with drifting or large movement artifacts were removed by visual inspection before analysis. Independent component analysis was applied to the data to remove the effects of blinks and eye movements. Remaining trials with EEG voltages exceeding 70 μV measured from peak to peak at any channel were also removed. After the application of the whole artifact correction–rejection procedure, a total of 12% of trials were rejected for the Go condition and 13% of trials for the NoGo condition.

TFR Analysis

For the computation of the time–frequency representation (TFR), spectral power (1–30 Hz) was obtained by convolving 6-cycle complex Morlet wavelets with each single-trial EEG epoch. The resulting EEG power representations were normalized by subtracting, in a frequency fashion, the baseline from the power in every time point and dividing this difference by the baseline mean power. The 500 ms preceding the onset of the polarity word (affirmative "sí," negative "no") was used as the baseline, which means that resulting TFRs reflect power changes relative to this period. Finally, before the statistical analysis, the single-trial TFRs were averaged separately for each of the eight experimental conditions.

The resulting averaged TFRs were evaluated statistically using the cluster-based random permutation method implemented in Fieldtrip (Maris and Oostenveld, 2007). This method deals with the multiple comparisons in frequency, space, and time by identifying, over the whole ERP segment (here, 61,500 points: 15 frequencies, from 1 to 30 Hz in two-frequency step, 100 temporal points, and 60 electrodes), clusters of significant differences between conditions (sample points in close frequency, spatial and temporal proximity) while effectively controlling for type 1 error. This statistical approach allows only for pairwise comparisons. Therefore, certain prior calculations were performed to evaluate the current experimental design.

First, for the main effect of Cue, sentence Polarity and Content were collapsed for each participant and Cue condition, and then a cluster-based randomization comparison was conducted on the resulting Go and NoGo TFRs. This strategy allowed us to identify clusters with significant inhibition-related effects. Next, the identified clusters were submitted to subsequent analyses using the whole experimental design. More specifically, for each participant ($n = 27$) and condition ($n = 8$), a single power value was obtained by averaging the frequency, temporal, and spatial points that formed the inhibition-related cluster, and further submitted to a three-way, repeated measures ANOVA with Cue (Go, NoGo), Polarity (Affirmative, Negative), and Content (Motor, Mental) as within-subject factors.

RESULTS

Behavioral Data

Go–NoGo Task

Table 3 shows the descriptive statistics for all the behavioral data. A Content (motor vs. mental) × Polarity (affirmative vs. negative) analysis of variance (ANOVA) was performed for Go reaction times (RT), after eliminating response errors (about 1.27%) and times exceeding 3 SD the individual mean (about 1.3%). The percentage of commission errors (in NoGo trials) and omission errors (in Go trials) were also computed and submitted to a Content × Polarity × Cue ANOVA. No significant effect was obtained for Go reaction times, $F(1, 26) = 1.17$, $\eta^2 = 0.043$. There were more commission than omission errors, but the effect did not reach the significant threshold, $F(1, 26) = 3.50$, $p = 0.07$, $\eta^2 = 0.119$. All the other effects showed F and η^2 values below 1.22 and 0.045, respectively.

Recognition Task

A Content × Polarity × Cue ANOVA was performed for response reaction times (RT), after eliminating response errors (about 1.7%) and times exceeding 3 SD the individual mean (about 3.7%). The main effect of Cue was significant, $F(1, 26) = 38.68$, $p < 0.001$, $\eta^2 = 0.598$, as responses were faster when preceded by Go (M = 1,488 ms) than by NoGo trials (M = 1,566 ms). Also, responses were faster for affirmative than for negative sentences, although this effect did not reach the significant threshold, $F(1, 26) = 2.94$, $p = 0.09$, $\eta^2 = 0.101$. There was no other significant effect for recognition latencies. Concerning the analysis on the proportion of errors, there was a significant main effect of Polarity, $F(1, 26) = 15.113$, $p < 0.001$, $\eta_p{}^2 = 0.36$, resulting from larger amount of errors in negative ($M = 0.06$) than affirmative ($M = 0.04$) sentences. Both the main effect of Cue, $F(1, 26) = 3.55$, $p = 0.07$,

TABLE 3 | Behavioral data.

		Motor		Mental	
		Polarity		Polarity	
Cue		Affirmative	Negative	Affirmative	Negative
Go/NoGo task					
Go	RT	357 (8.9)	361 (9.4)	362 (9.7)	362 (9.1)
	ERR	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
NoGo	RT				
	ERR	0.02 (0.03)	0.02 (0.03)	0.02(0.04)	0.02 (0.03)
Verification task					
Go	RT	1442 (57.3)	1494 (60.1)	1492 (55.4)	1525 (54.0)
	ERR	0.03 (0.01)	0.05 (0.01)	0.03 (0.01)	0.07 (0.01)
NoGo	RT	1546 (66.5)	1576 (60.6)	1571 (66.9)	1573 (57.0)
	ERR	0.06 (0.01)	0.06 (0.01)	0.04 (0.01)	0.06 (0.02)

Mean reaction times (RT) in milliseconds and error rates (0 to 1) in the Go/NoGo task and in the control task as a function of Content (motor vs. mental), Polarity (affirmative vs. negative), and Cue (Go vs. NoGo). The standard errors of the mean are shown in parentheses.

$\eta^2 = 0.120$, and the interaction between Cue and Polarity, $F(1, 26) = 2.38$, $\eta^2 = 0.084$, failed to reach significance. The F and η^2 values for all other effects were below 1.66 and 0.060, respectively.

TFRs Results: Inhibition-Related (NoGo vs. Go) Clusters

The time–frequency decomposition showed the expected pattern of strong increases in low frequency power (peaking around theta range, 4–7 Hz) after the cue signal onset, relative to the baseline period—the 500 ms preceding affirmative and negative particle onset. Power increases were maximal in fronto-central sites and larger for NoGo than for Go trials. Though of a small size, there were also power decreases for frequencies in the beta range (from 13 to 30 Hz), which were maximal in posterior regions but still visible in frontal and central sites, and stronger for Go than for NoGo trials. These inhibition-related differences in theta and beta power were part of the same and large cluster identified using the cluster-based method for the comparison between NoGo and Go trials, $T_{\mathrm{maxsum}} = 10{,}020$, $p < 0.001$. Thus, to better examine the dynamics within each frequency range, we conducted two additional cluster-based comparisons, one for the low-frequency (2–10 Hz) and another for the high-frequency (11–30 Hz) range, which we will describe below.

Theta Modulations

The cluster-based comparison for the low-frequency range (2–10 Hz) identified stronger power increase, relative to baseline, for NoGo than for Go trials, from approximately 200 to 650 ms after the cue onset $T_{\mathrm{maxsum}} = 1284$, $p < 0.001$. This cluster was maximal for the theta band (4–8 Hz) and located at medial sites of frontal and central regions. To explore the whole design in this inhibition-related cluster, a single cluster magnitude was computed, for each participant and condition, by averaging the amplitudes corresponding to the period (between 0.2 and 0.65 s), frequency (4–8 Hz), and topography showing maximal differences between NoGo and Go trials. We conducted next a Content (Motor, Mental) × Cue (Go, NoGo) × Polarity (Affirmative, Negative) ANOVA on this cluster magnitude. This analysis yielded the expected Cue main effect, $F(1, 26) = 18.55$, $p < 0.001$, $\eta^2 = 0.416$, and most important the interaction Cue × Polarity, $F(1, 26) = 9.76$, $p < 0.005$, $\eta^2 = 0.273$, but failed to produce main effects of Polarity, $F(1, 26) = 3.49$, $\eta^2 = 0.119$, or Content, $F(1, 26) < 1$, $\eta^2 = 0.007$, and of any interaction involving the Content Factor, $Fs\ (1, 26) < 1.76$, $\eta^2 < 0.070$. As **Figure 2** shows, theta power increases for the NoGo trials were smaller in the context of negative ($M = 0.58$) than in the context of affirmative sentences ($M = 0.70$), $t(26) = 3.07$, $p < 0.005$, Cohen's $d = 0.592$. In contrast, Go trials showed similar theta power magnitudes regardless of the polarity of the context, $t(26) = 1.17$, Cohen's $d = 0.225$ ($Ms = 0.34$ and 0.37 ms; $SEs = 0.04$). Moreover, although the inhibition-related effect—namely, stronger theta power increases for NoGo than Go trials—reached significance for the two polarity conditions, it was of a smaller size for the negative, $t(26) = 2.98$, $p < 0.001$, Cohen's $d = 0.574$, than for the affirmative context, $t(26) = 5.16$, $p < 0.001$, Cohen's $d = 0.994$. Thus, this pattern reflects minor increases in theta power for NoGo trials in the context of negative sentences, and hence confirms our prior findings (de Vega et al., 2016). Importantly, the factor Content had no effect on theta activity, which means that negation modulates theta band rhythms independently of the sentence content—either motor or mental.

Beta Modulations

As noted above, there was also a significant inhibition-related cluster in the higher frequency range (11–30 Hz), $T_{maxsum} = 9133$, $p < 0.001$. This reflects a larger decrease in beta power, relative to baseline, for Go than for NoGo trials in a set of right fronto-central electrodes, for the period between 210 and 370 ms after the cue onset and for frequencies ranging from 14 to 22 Hz (see **Figure 3**). The subsequent three-way ANOVA yielded effects of Cue, $F(1, 26) = 35.49$, $p < 0.001$, $\eta^2 = 0.577$, and the crucial interaction Cue × Polarity, $F(1, 26) = 24.05$, $p < 0.001$, $\eta^2 = 0.481$, but failed to show significant effects of Polarity, $F(1, 26) = 3.66$, $\eta^2 = 0.124$, Content, $F(1, 26) < 1$, $\eta^2 = 0.003$, and the interactions involving the Content factor, $Fs\ (1, 26) < 2.40$, $\eta^2 < 0.085$. In the context of affirmative sentences, NoGo trials showed smaller beta power decreases than Go trials, $t(26) = 8.26$, $p < 0.001$, Cohen's $d = 1.59$ ($Ms = -0.02$ and -0.18; $SEs = 0.02$), whereas in the context of negative sentences, there was no significant difference between NoGo ($M = -11$) and Go trials ($M = -14$), $t(26) = 1.68$, Cohen's $d = 0.325$. In addition, there were polarity effects for both Go and NoGo trials, but of opposite direction and distinct effect size. Hence, for Go trials, the power reduction was larger in the context of affirmative than in the context of negative polarity sentences, $t(26) = 2.33$, $p = 0.023$, Cohen's $d = 0.449$, while the reverse happened for NoGo trials: stronger decreases in negative than affirmative polarity sentences,

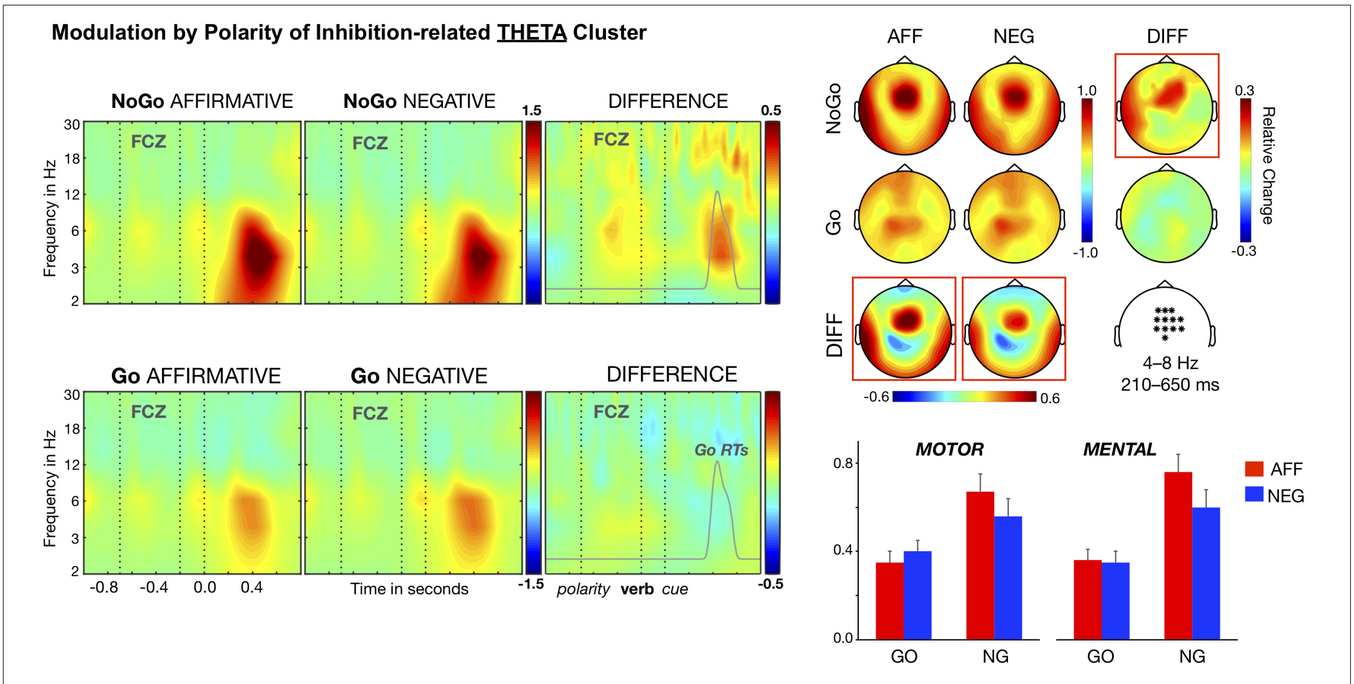


FIGURE 2 | Time–frequency analysis of the Polarity × Cue interaction. Theta band clusters (6–8 Hz) averaged over the fronto-central electrodes (marked in the white map) are shown in the left side panel. A statistically significant cluster of polarity difference arises in NoGo trials. Affirmative-NoGo trials elicited larger theta power than negative-NoGo trials in the time window of 210–650 ms after the cue onset. This NoGo cluster corresponds to the response inhibition stage, since it overlaps the distribution of Go RTs (gray curve). The distributions of theta band modulations on the scalp are shown in the upper right panel of the figure. The bars in the lower right panel show that the differential polarity effects in NoGo trials for the theta band was similar in motor and mental contents.

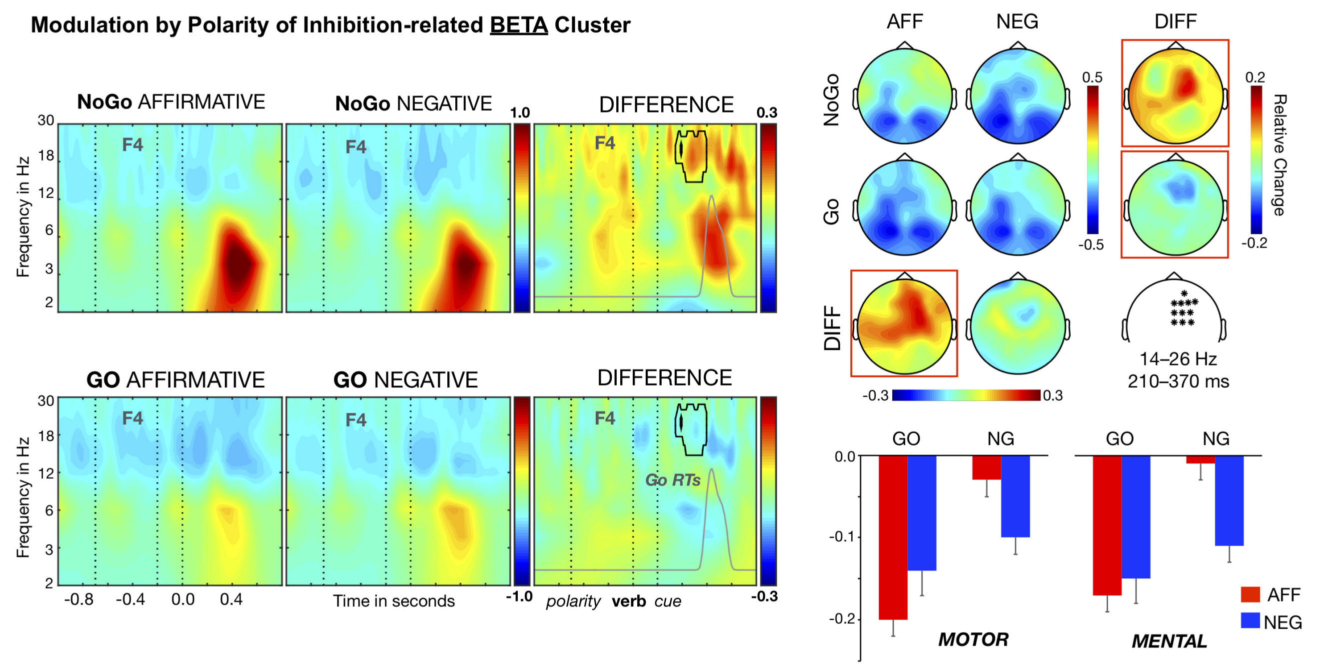


FIGURE 3 | Time–frequency analysis of the Polarity × Cue interaction. Beta band clusters (14–26 Hz) averaged over the right fronto-central electrodes (marked in the white map) are shown in the left side panel. A significant differential sentence polarity cluster (surrounded by a black line) arises in both NoGo and Go trials. Affirmative-NoGo trials elicited larger theta power than negative-NoGo trials in the time window of 210–370 ms after the cue onset, partially overlapping the theta modulation and likely corresponding to the response inhibition stage (see Go RT distribution, signaled as the gray curve). The distributions of beta band modulations on the scalp are shown in the upper right panel. The bars in the lower right panel show that the differential polarity effects in both NoGo and Go trials for the beta band were similar in motor and mental contents.

$t(26) = 4.58$, $p < 0.001$, Cohen's $d = 0.882$. Thus, like for the theta power cluster, negative sentences modulated the inhibition-related effect, mainly by interfering with the reduction of beta power elicited by the NoGo trials.

DISCUSSION

Recently, it has been reported that negation modulates some neurophysiological markers of inhibition, suggesting that neural inhibition mechanisms could be involved in the processing of sentential negation (de Vega et al., 2016; Papeo et al., 2016; Beltrán et al., 2018; Liu et al., unpublished). However, most of these studies were limited to action-related sentences and their conclusions cannot be generalized to other linguistic domains. In contrast, the current study aimed to test the inhibition hypothesis of negation with two types of sentences, referring either to motor actions or to abstract events. Like in a previous study (de Vega et al., 2016), a Go–NoGo task embedded in the comprehension of affirmative and negative sentences was used but adding the manipulation of sentence semantic content. As expected in inhibition trials (NoGo), the increase in fronto-central theta power—a well-known marker of inhibition processes—was larger in the context of affirmative than in the context of negative sentences, confirming previous results in the literature (de Vega et al., 2016). Most important, this interaction between polarity and response inhibition happened regardless of the type of negated content, suggesting that the response inhibition network operates as a content-free mechanism involved in the processing of negation.

de Vega et al. (2016) also reported a cue × polarity interactive modulation on delta rhythms (1 to 4 Hz) in their experiment 1, indexing a delayed post-response evaluation processes in Go trials. We did not replicate this delta modulation because the timing of the critical events in our trials (verb and cue presentation) was considerably faster in our study than in de Vega et al.'s experiment 1 and the Go theta effects require larger presentation times to emerge. Consistently, de Vega et al.'s experiment 2 employed the same event timing in trials as the current study and also did not find delta modulation. By contrast, we found a cue × polarity interaction on beta power oscillations over right fronto-central sites, which slightly precedes and overlaps the fronto-central theta effect, and therefore could be indexing the same inhibitory processes. In fact, fronto-central beta is also an accepted marker of response inhibition as reported elsewhere (Zhang et al., 2008; Huster et al., 2013; Wagner et al., 2018). The cue × polarity interaction on the beta band was driven by the strong differences in power between affirmative and negative sentences, especially in the context of inhibition (NoGo) trials. Again, there were no differential effects between the motor and the mental content on beta rhythm modulations, supporting the involvement of content-free inhibition associated with processing of sentential negation.

Concerning the behavioral measures, performance in the Go/NoGo task was characterized by a virtual ceiling effect, such that behavioral measures (Go reaction times and errors) were not sensitive to either the type of trial or the polarity of the sentence. Just like in the previous study (de Vega et al., 2016), the high accuracy rate in the current dual-task paradigm is likely due to the long inter-trial intervals between consecutive Go/NoGo cues, which precluded the setting of a very strong tendency to respond (e.g., Zamorano et al., 2014). Similarly, behavioral results confirmed the long-term effects of both cue and polarity on the sentence recognition task, by showing slower reactions for NoGo than Go trials, and higher error rates for negative than affirmative sentences; nonetheless, there was no significant interaction between the two factors, or of any of them with the type of negated content.

Theta and Beta Modulations

The increase in fronto-central theta band rhythms has been associated with inhibition-related processes in response inhibition tasks (e.g., Huster et al., 2013), and in this sense, our finding of content-free modulation by negation is consistent with the interpretation we advanced in previous studies (de Vega et al., 2016; Beltrán et al., 2018): that linguistic negation and response inhibition share inhibitory resources. Critically, the modulation by polarity of the inhibition effect over right-frontal beta power adds supporting evidence to this interpretation. Several studies, especially those using the SST, have already described that inhibition modulates oscillations in the beta band (e.g., Zhang et al., 2008; Wagner et al., 2018). More specifically, transient increases—i.e., synchronization—have been observed following the onset of the stop signal, which are either absent or reduced for non-inhibition (Go) and failed inhibition (stop) trials (e.g., Huster et al., 2013; Wagner et al., 2018). Furthermore, the use of electrocortical (ECoG) recordings indicates that beta synchronization originated at cortical areas around the right inferior frontal cortex (rIFC), an area strongly associated with the implementation of response inhibition (e.g., Swann et al., 2012; Aron et al., 2014). Our right-frontal beta effect shows the same pattern of differences between inhibition (NoGo trials) and non-inhibition trials (Go trials), as well as a similar distribution on the scalp, and therefore could be also interpreted as reflecting inhibition-related processes, which are modulated by sentence polarity.

The Generality of the Inhibitory Mechanism

The most important finding in the current study is that the interaction between negation and response inhibition signatures—i.e., fronto-central theta and right-frontal beta power—is equally modulated by motor and mental sentences. This finding considerably reinforces the hypothesis that the suppression effects of negation reported elsewhere (MacDonald and Just, 1989; Kaup and Zwaan, 2003; Kaup et al., 2006; Aravena et al., 2012; Bartoli et al., 2013) may be the consequence of applying a multipurpose inhibition mechanism to internal representations (de Vega et al., 2016; Beltrán et al., 2018; Papeo and de Vega, 2019). This proposal takes benefit from the idea of neural reuse, which holds that evolutionarily ancient mechanisms are redeployed to implement more recently acquired functions, while keeping the primary function (Anderson, 2010). This evolutionary strategy seems preferable because of being biologically less costly than developing *de novo* brain circuits. Our previous findings indicated that negation shares inhibitory mechanisms with response inhibition; however, they were limited regarding the generalizability of the effects, as they were obtained by combining a motor task (response inhibition in Go/NoGo or SST) with the comprehension of motor sentences. The current study extends these findings by probing that this interaction is not restricted to the negation of motor concepts, but it likely occurs regardless of the semantic modality of negated concepts. Thus, the reusing of inhibition networks—as indexed by modulations of some of its oscillatory markers—is a general characteristic of linguistic negation. However, note that the observed effect of negation, which is shared by action and non-action sentences, does not rule out that content-specific networks may also be affected differentially by the negation operator (Ghio et al., 2018). The general machinery of inhibitory control (indexed by our theta and beta modulations) could impact specific sensory-motor and semantic networks associated with particular contents, through specific cortico-cortical connections.

Inhibition and Control Monitoring

The embodied approach to language has highlighted what seems to be a clear case of neural reusing: the recruitment of action and perception brain systems for conceptual representations of meaning (e.g., for a review, Barsalou, 2016). However, the proper meaning of the negation markers, like the meaning of other grammatical and morphological elements in language, seems hard to ground on perceptual and motor systems. Our proposal offers an alternative way to account for negation processing from an embodied perspective; one in which negative operators recruit the processing systems involved in the regulation of other neural systems, including those of perceptions and actions. In other words, negation relies upon domain-general cognitive inhibition and/or control processes (de Vega et al., 2016; Dudschig and Kaup, 2018; Papeo and de Vega, 2019).

It is worth noting that response inhibition and control monitoring are two processes that, although related, could be functionally separated. In this sense, the modulation of theta oscillations has been clearly associated with response inhibition processes, involved in NoGo trials (in Go/NoGo tasks) or successful stop (in SST), and therefore could be considered a marker of neural inhibition (e.g., Huster et al., 2013). However, modulations in theta oscillations with source in medial prefrontal regions have also been reported in a variety of cognitive control tasks such as the Simon task or the flankers task, which demand conflict resolution and decision making rather than response inhibition (e.g., Nigbur et al., 2011; Cohen, 2014). The morphological and biological features of prefrontal neurons support oscillations in theta band, which could be associated with diverse high-order cognitive processes implemented in the same or neighbor populations of neurons in the prefrontal cortex (Cohen, 2014; Dippel et al., 2017). Accordingly, the finding of a modulation of fronto-central theta oscillations by negation is ambiguous, since it does not clearly specify whether it indexes response inhibition, control monitoring or both. In any case, the fact that negative sentences modulate theta oscillations in NoGo trials strongly supports that the mechanisms underlying response inhibition are involved to some extent in the semantics of negation.

Concerning the right fronto-central beta oscillations, it has been reported that they are selectively modulated in response inhibition tasks, supporting the claim that they constitute a genuine neurobiological marker of inhibitory processes (e.g., Zhang et al., 2008; Huster et al., 2013; Wagner et al., 2018). Consequently, the fact that right fronto-central beta oscillations are modulated by negation, especially in NoGo trials, could support a more specific interpretation of our results: sentential negation interacts with response inhibition processes, at least in the context of the Go–NoGo task. It remains to be tested whether similar modulations of beta oscillations rise for non-motor (cognitive) inhibitions. Our proposal is that negation conveys the recruitment of domain-general inhibitory control mechanisms, and hence it should interact with a variety of inhibition paradigms.

Note, however, that it could be possible that processing sentential negation recruits the two mentioned control mechanisms—response inhibition and conflict monitoring—at different moments (e.g., Giora et al., 2007; Orenes et al., 2014; Dudschig and Kaup, 2018). The two-step process assumes that negated information is activated first (e.g., *open door* for the sentence "*The door is not open*") and immediately followed by the updating of the alternative representation (e.g., *closed door*), which corresponds to the actual state of affairs (e.g., Kaup et al., 2006; Giora et al., 2007; Nieuwland and Kuperberg, 2008; Orenes et al., 2014; Tian and Breheny, 2016). The initial activation process is thought to be automatic, governed by memory-based associative operations, and very similar to that involved in the processing of affirmative sentences (Deutsch et al., 2006, 2009). By contrast, more controlled, rule-based processes are thought to intervene in the second step, inducing a change of the initial representation (Deutsch et al., 2009; Dudschig et al., 2019). Therefore, dealing with two opposing representations—the negated and the actual state of affairs—seems to demand to some extent conflict monitoring and selection of alternatives, followed by suppression or inhibition of the initial alternative (Dudschig and Kaup, 2018). Indeed, there is an interesting parallelism with the processing sequence involved in ordinary response inhibition tasks. In both the Go/NoGo and the SST, a strong tendency to respond is created, which conflicts with the inhibition cues (NoGo and Stop, respectively). This implies that, in inhibition trials, an initial step of response activation is followed by conflict detection—triggered by NoGo or Stop cues—and by the selection of an alternative course of action (inhibition plus updating).

Nonetheless, the current study probably gives an incomplete view of the neural dynamics in sentential negation processing. The interactive effects of negation on theta and beta rhythms were observed in a relatively early temporal snapshot, while the negation and the verb were still being integrated and at the time of the Go/NoGo cue presentation (this was also the case in previous studies with inhibition task paradigms: de Vega et al., 2016; Beltrán et al., 2018). At first sight, this timing seems inconsistent with the hypothetical two steps involved in the processing of negation. According to the two-step's proposal, the negation should not modulate the neural activity as early as we have found because the polarity marker supposedly does not initially affect the first-step representation. Even more, these early interactive effects could be compatible with a single-step model of negation, in which, for instance, inhibitory control processes could operate incrementally from the beginning precisely to impede the activation of the negated situation model. This might be especially plausible when the previous linguistic or pragmatic context biases the actual meaning of the negated sentence. However, it is also possible that this early interaction between inhibition and negation is reflecting neural processes that were not detected by previous studies reporting a late impact of negation on behavioral or ERP measures (e.g., for a recent review, Kaup and Dudschig, 2019). Most prior behavioral and ERP studies have mainly focused on detecting the representational states associated with negation by measuring indexes of semantic processing such as the N400 component in world-knowledge or semantic violation sentences (e.g., Fischler et al., 1983; Nieuwland and Kuperberg, 2008; Dudschig et al., 2018, 2019) or reaction times in probe recognition tasks (MacDonald and Just, 1989; Kaup, 2001; Kaup and Zwaan, 2003), neglecting the conflict monitoring and inhibition processes that underlie the two-step process dynamics. By contrast, the time–frequency analysis of the EEG in the context of the dual-task paradigm we employed here may reveal an early operation of the inhibitory control mechanism that governs the transition from the initial to the final representation of the negated events (Dudschig et al., 2019) or prevents the representation of the negated situation. Interesting issues for further research are the extent to which the early inhibitory control process is automatically activated by negation operators, and the role that pragmatics (e.g., the pre-activation of negated meaning by the preceding context) plays in the initiation of conflict monitoring and the inhibition processes associated with negation.

This study has some limitations that must be overcome in further research. First, although the modulation of theta and beta signatures by negation was strong and content-free, we may note that it was obtained in a dual-task paradigm. Namely, the modulatory effects only emerged in the context of inhibitory NoGo trials. Future experiments, with alternative techniques (neuroimaging, TMS, electrophysiological functional localizers), will be needed to obtain differential neurobiological signatures of inhibition and control during the comprehension of sentences differing in polarity, without performing any other parallel task. Second, although the study made an important step in generalizing the hypothesis of neural inhibition, negation has many semantic and pragmatic dimensions, and the generality of the hypothesis needs additional proofs employing more diverse contents, such as perceptual, existential, or emotional. Third, the imperative format of our negative sentences can be partially responsible for the observed effects, given the fact that imperatives are functionally equivalent to a stop signal (de Vega et al., 2016). Additional studies are needed, using purely declarative sentential negations to determine whether inhibitory control processes also underlie other syntactic structures. Fourth, a more careful evaluation is needed to explore the relative role played by inhibition and control monitoring mechanisms during the processing of the two alternative meanings of negation. Fifth, our recording of interactive effects was constrained to an early time window associated with the verb, and other possible effects beyond this point were not registered in the study. In principle, integrative processes of negation considerably extend in time and we did not exhaust the analysis of potentially relevant effects, for instance of the content. What we registered was the early impact of the general control inhibitory mechanisms in negation, which could spread out to content-specific networks in a later stage. In other words, this inhibitory control mechanism may govern the deactivation of content-specific networks, such as the motor system, the visual system or the semantic hubs for action, visual or abstract sentences, respectively. The study of the functional links between the general inhibitory system and the content-specific networks during the processing of sentential negation is a relevant topic for further research.

CONCLUSION

In this article, we obtained two convergent electrophysiological signatures (theta and beta oscillations) confirming that sentential negation may share neural processes with response inhibition. We propose that this inhibitory mechanism contributes to fill a gap in the current models of negation processing. The two-step account proposes dynamic changes during the processing of negation: from an initial representation of the negated state of affairs to its suppression and replacement by a representation of the actual state of affairs. The incremental single-stage model posits that the representation of the factual meaning of negation is immediately activated, while the alternative (negated situation) representation is blocked. However, none of these models provide any mechanism responsibly for governing the suppression or blocking of the negated situation. We propose that our data support the idea that the neural network of inhibition is a plausible mechanism, either to impede the activation of the negated events representation (in the single-step model) or to produce the transition between the initial (negated events) representation and the final (current events) representation in the two-step process. This mechanism works in a content-free manner, given the fact that the same neurobiological markers were equally sensitive to negative motor sentences and to negative mental sentences.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comité de Ética de la Investigación y Bienestar Animal, Vicerrectorado de Investigación y Transferencia de Conocimiento, Universidad de La Laguna. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DB: experimental design, data analysis, and manuscript writing. YM: linguistic materials, experimental script data collection, and manuscript writing. EG-M: data collection and manuscript writing. MdV: experimental design, experimental materials, planning of data analysis, and manuscript writing.

FUNDING

This work was funded by the Spanish MINECO (Grant PSI2015-66277-R to MdV), the Cabildo Insular de Tenerife, the COP of Las Palmas de Gran Canaria, and the European Regional Development Funds.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01782/full#supplementary-material

REFERENCES

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266. doi: 10.1017/s0140525x10000853

Aravena, P., Delevoye-Turrell, Y., Deprez, V., Cheylus, A., Paulignan, Y., Frak, V., et al. (2012). Grip force reveals the context sensitivity of language-induced motor activity during "action words" processing: evidence from sentential negation. *PLoS One* 7:e50287. doi: 10.1371/journal.pone.0050287

Aravena, P., Hurtado, E., Riveros, R., Cardona, J. F., Manes, F., and Ibáñez, A. (2010). Applauding with closed hands: neural signature of action-sentence compatibility effects. *PLoS One* 5:e11751. doi: 10.1371/journal.pone.0011751

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn. Sci.* 18, 177–185. doi: 10.1016/j.tics.2013.12.003

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychon. Bull. Rev.* 23, 1122–1142. doi: 10.3758/s13423-016-1028-3

Bartoli, E., Tettamanti, A., Farronato, P., Caporizzo, A., Moro, A., Gatti, R., et al. (2013). The disembodiment effect of negation: negating action-related sentences attenuates their interference on congruent upper limb movements. *J. Neurophysiol.* 109, 1782–1792. doi: 10.1152/jn.00894.2012

Beltrán, D., Muñetón-Ayala, M., and de Vega, M. (2018). Sentential negation modulates inhibition in a stop-signal task. Evidence from behavioral and ERP data. *Neuropsychologia* 112, 10–18. doi: 10.1016/j.neuropsychologia.2018.03.004

Beltrán, D., Orenes, I., and Santamaría, C. (2008). Context effects on the spontaneous production of negation. *Intercult. Pragmat.* 5, 409–419. doi: 10.1515/IPRG.2008.020

Bokura, H., Yamaguchi, S., and Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clin. Neurophysiol.* 112, 2224–2232. doi: 10.1016/S1388-2457(01)00691-5

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., and Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia* 46, 743–756. doi: 10.1016/j.neuropsychologia.2007.10.007

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cogn. Brain Res.* 24, 355–363. doi: 10.1016/j.cogbrainres.2005.02.020

Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends Neurosci.* 37, 480–490. doi: 10.1016/j.tins.2014.06.004

Dale, R., and Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cogn. Sci.* 35, 983–996. doi: 10.1111/j.1551-6709.2010.01164.x

de Vega, M., León, I., Hernández, J. A., Valdés, M., Padrón, I., and Ferstl, E. C. (2014). Action sentences activate sensory-motor regions in the brain independently of their status of reality. *J. Cogn. Neurosci.* 26, 1363–1376. doi: 10.1162/jocn_a_00559

de Vega, M., Morera, Y., León, I., Beltrán, D., Casado, P., and Martín-Loeches, M. (2016). Sentential negation might share neurophysiological mechanisms with action inhibition. Evidence from frontal theta rhythm. *J. Neurosci.* 36, 6002–6010. doi: 10.1523/JNEUROSCI.3736-15.2016

Deutsch, R., Gawronski, B., and Strack, F. (2006). At the boundaries of automaticity: negation as reflective operation. *J. Pers. Soc. Psychol.* 91:385. doi: 10.1037/0022-3514.91.3.385

Deutsch, R., Kordts-Freudinger, R., Gawronski, B., and Strack, F. (2009). Fast and fragile: a new look at the automaticity of negation processing. *Exp. Psychol.* 56, 434–446. doi: 10.1027/1618-3169.56.6.434

Dippel, G., Muckschel, M., Ziemssen, T., and Bestle, C. (2017). Demands on response inhibition processes determine modulations of theta band activity in superior frontal areas and correlations with pupillometry—Implications for the norepinephrine system during inhibitory control. *Neuroimage* 157, 575–585. doi: 10.1016/j.neuroimage.2017.06.037

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., and Carreiras, M. (2013). EsPal: one-stop shopping for spanish word properties. *Behav. Res. Methods* 45, 1246–1258. doi: 10.3758/s13428-013-0326-1

Dudschig, C., and Kaup, B. (2018). How does not "left" become "right"? Electrophysiological evidence for a dynamic conflict-bound negation processing account. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 716–728. doi: 10.1037/xhp0000481

Dudschig, C., Mackenzie, I. G., Leuthold, H., and Kaup, B. (2018). Environmental sound priming: does negation modify N400 cross-modal priming effects? *Psychonomic Bull. Rev.* 25, 1441–1448. doi: 10.3758/s13423-017-1388-3

Dudschig, C., Mackenzie, I. G., Maienborn, C., Kaup, B., and Leuthold, H. (2019). Negation and the N400: investigating temporal aspects of negation integration using semantic and world-knowledge violations. *Lang. Cogn. Neurosci.* 34, 309–319. doi: 10.1080/23273798.2018.1535127

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., and Perry, N. W. Jr. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology* 20, 400–409. doi: 10.1111/j.1469-8986.1983.tb00920.x

Foroni, F., and Semin, G. R. (2013). Comprehension of action negation involves inhibitory simulation. *Front. Human Neurosci.* 7:209. doi: 10.3389/fnhum.2013.00209

García, A. M., and Ibáñez, A. (2016). A touch with words: dynamic synergies between manual actions and language. *Neurosci. Biobehav. Rev.* 68, 59–95. doi: 10.1016/j.neubiorev.2016.04.022

García-Marco, E., Morera, Y., Beltrán, D., de Vega, M., Herrera, E., Sedeño, L., et al. (2019). Negation markers inhibit motor routines during typing of manual action verbs. *Cognition* 182, 286–293. doi: 10.1016/j.cognition.2018.10.020

Ghio, M., Haegert, K., Vaghi, M. M., and Tettamanti, M. (2018). Sentential negation of abstract and concrete conceptual categories: a brain decoding multivariate pattern analysis study. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373, 20170124. doi: 10.1098/rstb.2017.0124

Giora, R., Fein, O., Aschkenazi, K., and Alkabets-Zlozover, I. (2007). Negation in context: a functional approach to suppression. *Discourse Process.* 43, 153–172. doi: 10.1080/01638530709336896

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565. doi: 10.3758/BF03196313

Harper, J., Malone, S. M., and Bernat, E. M. (2014). Theta and delta band activity explain N2 and P3 ERP component activity in a go/no-go task. *Clin. Neurophysiol.* 125, 124–132. doi: 10.1016/j.clinph.2013.06.025

Herrera, E., Cuetos, F., and Ribacoba, R. (2012). Verbal fluency in Parkinson's disease patients on/off dopamine medication. *Neuropsychologia* 50, 3636–3640. doi: 10.1016/j.cortex.2010.12.007

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., and Herrmann, C. S. (2013). Electroencephalography of response inhibition tasks: functional networks and cognitive contributions. *Int. J. Psychophysiol.* 87, 217–233. doi: 10.1016/j.ijpsycho.2012.08.001

Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Mem. Cognit.* 29, 960–967. doi: 10.3758/bf03195758

Kaup, B., and Dudschig, C. (2019). "Understanding negation. Issues on the processing of negation," in *The Oxford Handbook on Negation*, eds V. Deprez and M. T. Espinal.

Kaup, B., Lüdtke, J., and Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: is a door that is not open mentally closed? *J. Pragmat.* 38, 1033–1050. doi: 10.1016/j.pragma.2005.09.012

Kaup, B., and Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *J. Exp. Psychol. Learn. Mem. Cogn.* 29:439. doi: 10.1037/0278-7393.29.3.439

Khemlani, S., Orenes, I., and Johnson-Laird, P. N. (2012). Negation: a theory of its meaning, representation, and use. *J. Cogn. Psychol.* 24, 541–559. doi: 10.1080/20445911.2012.660913

Klepp, A., Niccolai, V., Buccino, G., Schnitzler, A., and Biermann-Ruben, K. (2015). Language motor interference reflected in MEG beta oscillations. *Neuroimage* 109, 438–448. doi: 10.1016/j.neuroimage.2014.12.077

Liuzza, M. T., Candidi, M., and Aglioti, S. M. (2011). Do not resonate with actions: sentence polarity modulates cortico-spinal excitability during action-related sentence reading. *PLoS One* 6:e16855. doi: 10.1371/journal.pone.0016855

MacDonald, M. C., and Just, M. A. (1989). Changes in activation levels with negation. *J. Exp. Psychol. Learn. Mem. Cogn.* 15:633. doi: 10.1037/0278-7393.15.4.633

Maguire, M. J., Brier, M. R., and Ferree, T. C. (2010). EEG theta and alpha responses reveal qualitative differences in processing taxonomic versus thematic semantic relationships. *Brain Lang.* 114, 16–25. doi: 10.1016/j.bandl.2010.03.005

Maris, E., and Oostenveld, D. R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Mayo, R., Schul, Y., and Burnstein, E. (2004). "I am not guilty" vs "I am innocent": successful negation may depend on the schema used for its encoding. *J. Exp. Soc. Psychol.* 40, 433–449. doi: 10.1016/j.jesp.2003.07.008

Moody, C. L., and Gennari, S. P. (2010). Effects of implied physical effort in sensory-motor and pre-frontal cortex during language comprehension. *Neuroimage* 49, 782–793. doi: 10.1016/j.neuroimage.2009.07.065

Moreno, I., de Vega, M., and León, I. (2013). Understanding action language modulates oscillatory mu and beta rhythms in the same way as observing actions. *Brain Cogn.* 82, 236–242. doi: 10.1016/j.bandc.2013.04.010

Moreno, I., De Vega, M., León, I., Bastiaansen, M., Lewis, A. G., and Magyari, L. (2015). Brain dynamics in the comprehension of action-related language. A time–frequency analysis of mu rhythms. *Neuroimage* 109, 50–62. doi: 10.1016/j.neuroimage.2015.01.018

Nieuwland, M. S., and Kuperberg, G. R. (2008). When the truth is not too hard to handle: an event-related potential study on the pragmatics of negation. *Psychol. Sci.* 19, 1213–1218. doi: 10.1111/j.1467-9280.2008.02226.x

Nigbur, R., Ivanova, G., and Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clin. Neurophysiol.* 122, 2185–2194. doi: 10.1016/j.clinph.2011.03.030

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci*. 2011:156869. doi: 10.1155/2011/156869

Orenes, I., Beltrán, D., and Santamaría, C. (2014). How negation is understood: evidence from the visual world paradigm. *J. Mem. Lang.* 74, 36–45. doi: 10.1016/j.jml.2014.04.001

Orenes, I., Moxey, L., Scheepers, C., and Santamaría, C. (2016). Negation in context: evidence from the visual world paradigm. *Q. J. Exp. Psychol. A* 69, 1082–1092. doi: 10.1080/17470218.2015.1063675

Papeo, L., Hochmann, J. R., and Battelli, L. (2016). The default computation of negated meanings. *J. Cogn. Neurosci.* 28, 1980–1986. doi: 10.1162/jocn_a_01016

Papeo, L., Vallesi, A., Isaja, A., and Rumiati, R. I. (2009). Effects of TMS on different stages of motor and non-motor verb processing in the primary motor cortex. *PLoS One* 4:e4508. doi: 10.1371/journal.pone.0004508

Papeo, L., and de Vega, M. (2019). "The neurobiology of sentential negation: the hypothesis of the functional integration of language and inhibition," in *The Oxford Handbook on Negation*, eds V. Deprez and M. T. Espinal.

Swann, N. C., Cai, W., Conner, C. R., Pieters, T. A., Claffey, M. P., George, J. S., et al. (2012). Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *Neuroimage* 59, 2860–2870. doi: 10.1016/j.neuroimage.2011.09.049

Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., et al. (2008). Negation in the brain: modulating action representations. *Neuroimage* 43, 358–367. doi: 10.1016/j.neuroimage.2008.08.004

Tian, Y., and Breheny, R. (2016). "Dynamic pragmatic view of negation processing," in *Negation and polarity: Experimental perspectives*, eds P. Larrivée and C. Lee (Cham: Springer), 21–43. doi: 10.1007/978-3-319-17464-8_2

Tomasino, B., Fink, G. R., Sparing, R., Dafotakis, M., and Weiss, P. H. (2008). Action verbs and the primary motor cortex: a comparative TMS study of silent reading, frequency judgments, and motor imagery. *Neuropsychologia* 46, 1915–1926. doi: 10.1016/j.neuropsychologia.2008.01.015

Tomasino, B., Weiss, P. H., and Fink, G. R. (2010). To move or not to move: imperatives modulate action-related verb processing in the motor system. *Neuroscience* 169, 246–258. doi: 10.1016/j.neuroscience.2010.04.039

van Elk, M., van Schie, H. T., Zwaan, R. A., and Bekkering, H. (2010). The functional role of motor activation in language processing: motor cortical oscillations support lexical-semantic retrieval. *Neuroimage* 50, 665–677. doi: 10.1016/j.neuroimage.2009.12.123

Wagner, J., Wessel, J. R., Ghahremani, A., and Aron, A. R. (2018). Establishing a right frontal beta signature for stopping action in scalp EEG: implications for testing inhibitory control in other task contexts. *J. Cogn. Neurosci.* 30, 107–118. doi: 10.1162/jocn_a_01183

Wason, P. C., and Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*, Vol. 86. Cambridge, MA: Harvard University Press.

Zamorano, F., Billeke, P., Hurtado, J. M., López, V., Carrasco, X., Ossandón, T., et al. (2014). Temporal constraints of behavioral inhibition: relevance of inter-stimulus interval in a Go–Nogo task. *PLoS One* 9:e87232. doi: 10.1371/journal.pone.0087232

Zhang, Y., Chen, Y., Bressler, S. L., and Ding, M. (2008). Response preparation and inhibition: the role of the cortical sensorimotor beta rhythm. *Neuroscience* 156, 238–246. doi: 10.1016/j.neuroscience.2008.06.061

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 August 2019
doi: 10.3389/fpsyg.2019.01985

Priming Effects of Focus in Mandarin Chinese

Mengzhu Yan and Sasha Calhoun*

School of Linguistics and Applied Language Studies, Victoria University of Wellington, Wellington, New Zealand

Psycholinguistic research has long established that focus-marked words have a processing advantage over other words in an utterance, e.g., they are recognized more quickly and remembered better. More recently, studies have shown that listeners infer contextual alternatives to a focused word in a spoken utterance, when marked with a contrastive accent, even when the alternatives are not explicitly mentioned in the discourse. This has been shown by strengthened priming of contextual alternatives to the word, but not other non-contrastive semantic associates, when it is contrastively accented, e.g., after hearing "The *customer* opened the window," *salesman* is strongly primed, but not *product*. This is consistent with Rooth's (1992) theory that focus-marking signals the presence of alternatives to the focus. However, almost all of the research carried out in this area has been on Germanic languages. Further, most of this work has looked only at one kind of focus-marking, by contrastive accenting (prosody). This paper reports on a cross-modal lexical priming study in Mandarin Chinese, looking at whether focus-marking heightens activation, i.e., priming, of words and their alternatives. Two kinds of focus-marking were investigated: prosodic and syntactic. Prosodic prominence is an important means of focus-marking in Chinese, however, it is realized through pitch range expansion, rather than accenting. The results showed that focused words, as well as their alternatives, were primed when the subject prime word carried contrastive prosodic prominence. Syntactic focus-marking, however, did not enhance priming of focused words or their alternatives. Non-contrastive semantic associates were not primed with either kind of focus-marking. These results extend previous findings on focus and alternative priming for the first time to Chinese. They also suggest that the processing advantages of focus, including priming alternatives, are particularly related to prosodic prominence, at least in Chinese and Germanic languages. This research sheds light on what linguistic mechanisms listeners use to identify important information, generate alternatives, and understand implicature necessary for successful communication.

Keywords: alternatives, contrast, focus, prosody, syntax, Mandarin Chinese

OPEN ACCESS

Edited by:
Sophie Repp,
University of Cologne, Germany

Reviewed by:
Ingo Feldhausen,
Goethe University Frankfurt, Germany
Scott Fraundorf,
University of Pittsburgh, United States

****Correspondence:***
Mengzhu Yan
mengzhu.yan@vuw.ac.nz

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *23 December 2018*
Accepted: *13 August 2019*
Published: *30 August 2019*

Citation:
Yan M and Calhoun S (2019) Priming Effects of Focus in Mandarin Chinese. Front. Psychol. 10:1985. doi: 10.3389/fpsyg.2019.01985

1. INTRODUCTION

The process of successful comprehension in spoken discourse involves more than understanding the words that are said. Listeners need to attend most carefully to the part of an utterance which gives the most important information, that which updates the common ground. Further, as the theme of this research topic attests, they frequently need to infer information which is not directly expressed in the utterance they are listening to, such as alternatives to one of the elements in the utterance. Focus-marking, e.g., by contrastive accenting, allows listeners to do this. For example, when a speaker says "The *customer* closed the window" (italics indicate a contrastive accent), this implies not simply that the customer closed the window, but also that *the customer* is the important information which updates an explicit or implicit 'question-under-discussion' (QUD) like "Who closed the window?" (Roberts, 1996), and that it is relevant that someone else, e.g., *the salesman*, could have closed the window. To make the communication successful, listeners must be able to successfully identify the focus, and thus infer the alternatives intended by the speaker even when these are not available in the context. Psycholinguistic studies since the 1970s have shown that focused words are indeed attended to more than defocused or unfocused words: they are recognized faster and remembered better (e.g., Cutler and Fodor, 1979; Birch and Garnsey, 1995; Cutler et al., 1997; Birch et al., 2000; Akker and Cutler, 2003) More recently, there has been mounting evidence that listeners activate alternatives in sentences like these, even when the alternatives are not explicitly mentioned in the discourse; and this activation is facilitated by contrastive accenting (e.g., Braun and Tagliapietra, 2010; Gotzner et al., 2016; Husband and Ferreira, 2016).

Focus-marking therefore has at least two key functions: to indicate the information which updates the common ground, and, following the alternative semantics theory proposed by Rooth (1992), to indicate contextually-relevant alternatives. There are a number of different linguistic means to indicate focus, including contrastive accenting (or prominence), certain syntactic constructions, e.g., clefts, and morphological markers (Féry and Ishihara, 2016). However, most of the psycholinguistic work in this area has concentrated on contrastive accenting. While some work has shown that clefting strengthens attention and memory for focused words (Birch and Garnsey, 1995; Birch et al., 2000; Kember et al., 2016a,b), to our knowledge, no previous studies have investigated whether other focus-marking mechanisms also activate alternatives, e.g., clefting, in the absence of prosodic focus-marking. Across languages, morphosyntactic means of marking focus are as common as prosodic, and if focus is the underlying mechanism this should be the case. However, if the activation of alternatives is rather related to prosodic prominence, which enhances the salience of the prominent word, and therefore its processing, these other focus-marking mechanisms would not activate alternatives.

Further, to our knowledge, all of the studies in this area have been carried out on Germanic languages, which have very similar prosodic systems. In this paper, we report on a cross-modal lexical priming study carried out in Mandarin Chinese (hereafter Chinese). Prosodic prominence is a key marker of focus in Chinese, however, prominence in Chinese is marked differently from Germanic languages, through pitch register expansion rather than pitch accenting (Xu, 1999). Focus can also be marked by cleft constructions in Chinese (Fang, 1995; Paul and Whitman, 2008). The study therefore expands the cross-linguistic validity of the effects. The study looks at priming of subject arguments in spoken sentences; looking at whether subject words and their alternatives are primed by syntactic as well as prosodic cues to focus in Chinese.

In section 2, we will define focus, drawing on the theoretical literature, and present the prosodic and syntactic markers of focus in Chinese explored in this study. We then review previous studies regarding the effect of focus on speech processing, including the recent research on the role of focus in priming alternatives.

2. FOCUS AND FOCUS-MARKING

Focus is a key part of information structure. During a discourse, speakers build a *common ground* of propositions relevant to the context they believe to be established with the other speaker(s) (Stalnaker, 1974; Clark, 1996). To facilitate this, each utterance has an *information structure*, i.e., each argument, predicate, etc. is marked as to how it refers back to, alters and/or updates the common ground (Chafe, 1976; Féry and Krifka, 2008; Krifka, 2008). One key kind of marking is *focus-marking*. There are two main definitions of focus, which are in principle orthogonal to each other (Calhoun, 2010; Vallduví, 2016). Under the first the focus, or rheme, is the part of the utterance which updates the common ground, or is new in relation to the current question-under-discussion (QUD) (see e.g., Ginzburg, 1994; Roberts, 1996; Vallduví, 2016). We will call this QUD-focus. Under the second, the focus, or contrast, indicates "the presence of alternatives that are relevant for the interpretation of linguistic expressions" (Rooth, 1992; Krifka, 2008, p. 247). We will call this contrastive focus. Both of these are illustrated in the following (bold indicates the prosodic prominence, F shows the focus):[1]

(1) a. 展厅的一个窗户大敞着，每个人都觉得冷。
"There was a window wide open in the showroom, making everyone cold."

b. [**顾客**]$_F$ 关上 了 窗户。
gu4ke4 guan1 shang4 le0 chuang1hu0
customer close PRF window
"[The **customer**]$_F$ (has) closed the window."

c. #顾客 关上 了 [**窗户**]$_F$。
gu4ke4 guan1shang4 le0 **chuang1hu0**
customer close PRF **window**
#"The customer (has) closed [the **window**]$_F$."

In (1b), 顾客 "the customer" is the QUD-focus as it updates the common ground, giving new information on 窗户"the window," which was mentioned in (1a), i.e., there is an implied question of 谁关上了窗户? "Who closed the window?" The

[1]The first tier of (1b) indicates the Chinese characters. The second tier shows Pinyin, which is the official romanisation of Chinese. The numbers (0–4) represent tones (neutral, high, rising, low, falling). In the third tier, the following abbreviations are used in glosses: PRF = perfective aspect, COP, copula; S, subject; V, verb; O, object. 的 (*DE*) is glossed as *DE* following the current literature such as Paul and Whitman (2008) and Hole (2011). Note that Mandarin Chinese 的 (*DE*) has multiple uses (apart from its association with past tense reading in this paper), which includes its function as a complementizer, a nominalizer and others (see e.g., Paul and Whitman, 2008; Xie, 2012).

focus is marked by prosodic prominence on 顾客 "customer." This prominence can also indicate focus according to the second definition, i.e., contrastive focus, implying a contextually-appropriate set of alternatives to 顾客 "customer," e.g., 店员 "salesman." Non-contrastive associates, i.e., words that are semantically associated with 顾客 "customer," but cannot replace it in the sentence, e.g., 产品 "product," are not in the alternative set. Likewise, alternatives to the unfocused argument, i.e., 窗户 "window," are not implied. As can be seen in this example, while QUD-focus and contrastive focus are in principle separable, in practice the same constituent in a sentence is often focused by either definition. When the sentence is said with prosodic prominence on 窗户 "window," as in (1c), this is incongruent, as focus on the object does not match the context by either focus definition: an implied question of 顾客关上了什么? "What did the customer close?" is odd as 顾客 "the customer" was not mentioned; likewise, alternatives to 窗户 "window" are odd as only the window is mentioned as being open.

In Chinese, prosodic prominence is a key marker of focus (Xu, 1999; Wang and Xu, 2006; Chen and Gussenhoven, 2008). Prosodic prominence is not realized by pitch accenting, as in Germanic languages, as the lexical tone determines the local F0 curve of each syllable. Rather, prosodic prominence is realized through pitch register. The pitch range in the focused word is expanded, and the region following the focus compressed (e.g., Xu, 1999; Wang and Xu, 2006; Chen and Gussenhoven, 2008) (see **Figure 1**). The focused word is also realized with longer duration and higher mean intensity (e.g., Xu, 1999; Chen and Gussenhoven, 2008; Chen et al., 2009). When the focus is on the subject, the following pitch range is heavily reduced (as in Germanic languages). When the focus is on the final object, which is the default position for primary prominence in Chinese, the pitch range in the pre-focal region is still relatively wide (again, similar to Germanic languages).

As mentioned above, prosodic prominence is not the only means of marking focus. Across languages, there are a number of other cues that mark focus including morphosyntactic

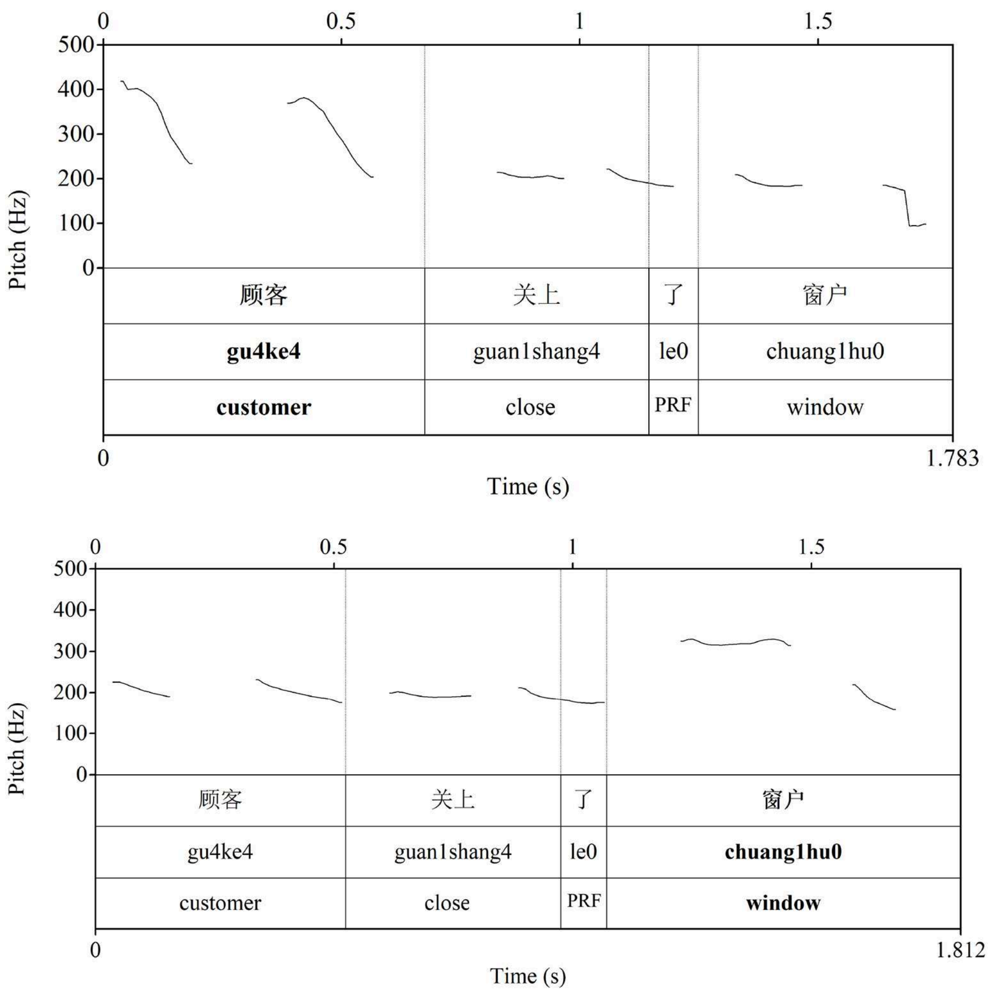


FIGURE 1 | A comparison of primary prosodic prominence on the subject **(Top)** and object **(Bottom)** in Chinese (see text for details).

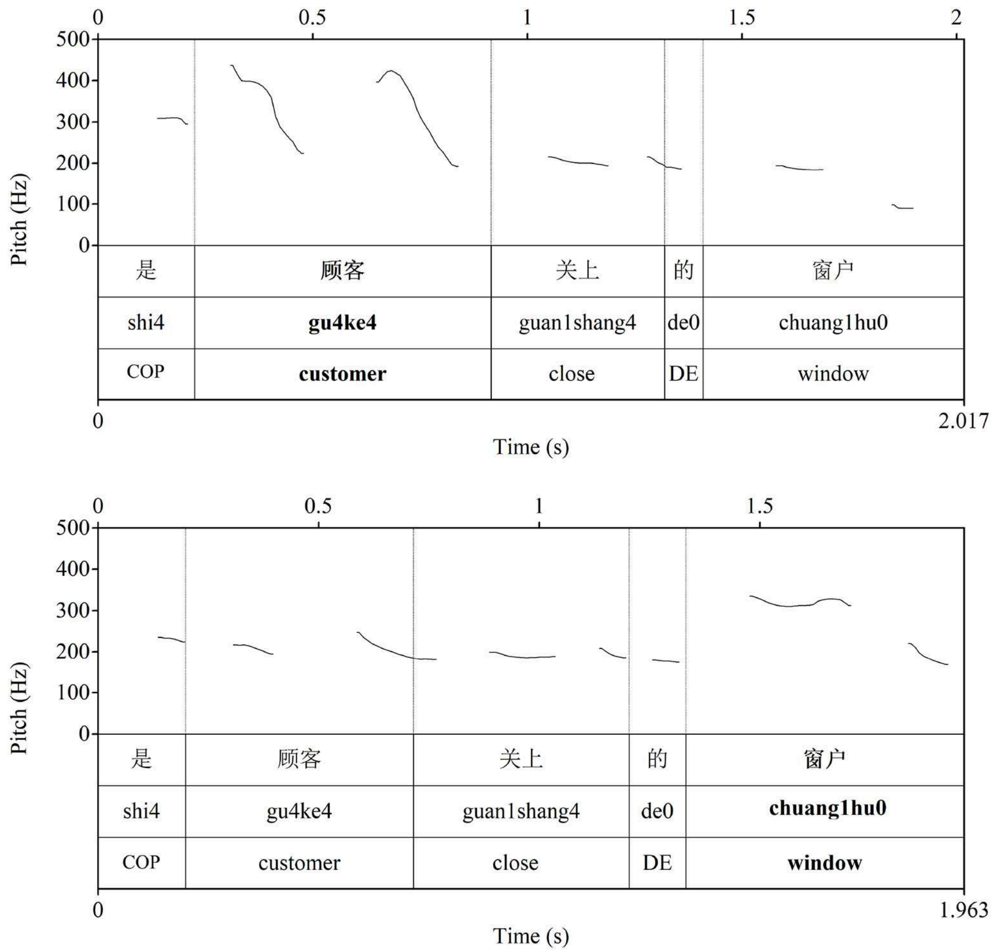


FIGURE 2 | Examples of the prosody of an ScleftS **(Top)** and ScleftO **(Bottom)** sentence in Chinese.

cues (e.g., clefts) and focus particles (e.g., *only*, *even*) (Féry and Ishihara, 2016). For instance, in Chinese, like in many languages, clefts can mark focus (Lambrecht, 2001; Paul and Whitman, 2008), as in the following:

(2) 是 [**顾客**]$_F$ 关上 的 窗户
shi4 **gu4ke4** guan1shang4 de0 chuang1hu0
COP **customer** close DE window
"It was [the **customer**]$_F$ who closed the window."
(COP S V DE O)

(3) 顾客 是 关上 的 [**窗户**]$_F$
gu4ke4 shi4 guan1shang4 de0 **chuang1hu0**
customer COP close DE **window**
"It was [the **window**]$_F$ that the customer closed."
(S COP V DE O)

Clefts are marked morphosyntactically in Chinese using the 是...的 (*SHI...DE*) construction, without changing the word order (Fang, 1995). For instance, for subject focus, as in (2), the copula 是 (*SHI*) occurs immediately before the subject, and 的 (*DE*) either before or after the object. When 的 (*DE*) appears before the object, the sentence is past tense (Hole, 2011). In this paper, we use the pre-object 的 (*DE*) (see Simpson and Wu, 2002; Paul and Whitman, 2008; Hole, 2011 for an overview of the SHI...DE cleft construction). For object focus, as in (3), the copula 是 (*SHI*) occurs before the verb, and the 是...的 (*SHI...DE*) construction does not change the word order. (2) marks focus on the subject, and like (1b) would be compatible with the context in (1a), while (3) marks object focus and would not be coherent following (1a).

The prosodic prominence normally falls on the cleft head, as shown in this example (see the top example in **Figure 2**). In (2), the cleft marks both QUD-focus on the cleft head, and contrastive focus, implying alternatives to it (Fang, 1995; É Kiss, 1998; Lambrecht, 2001). Further, it has been claimed that clefts have an exhaustive implication that focus-marking with prosodic prominence does not necessarily have (É Kiss, 1998; Krifka, 2008). The cleft rules out other alternatives in the context of that proposition. For example, it would be possible after (1b) to continue 而且店员帮助了她 "and the salesman helped her," but this would be not possible, or pragmatically odd, after (2).

While the prosodic prominence normally falls on the cleft head, it can also fall in the main clause (see the bottom example in **Figure 2**). In these cases, the QUD-focus is usually analyzed as being in the main clause, i.e., cued by the prosodic prominence (Prince, 1978; Delin and Oberlander, 1995; Lambrecht, 2001; Hole, 2011; Hedberg, 2013; Feldhausen and Vanrell, 2015), as in the following (note in both Chinese and English it is possible to have a secondary prominence, or accent, on the cleft head, however, the nuclear prominence is in the main clause):

(4) a. 展厅的门窗大敞着，店员和顾客看到每个人都很冷。
"The window and the door were wide open in the showroom. The salesman and the customer could see everyone was cold."

b. 是 [顾客]$_F$ 关上 的 [**窗户**]$_F$
shi4 gu4ke4 guan1shang4 de0 **chuang1hu0**
COP customer close DE **window**
"It was [the customer]$_F$ who closed [the **window**]$_F$."

c. 是 [店员]$_F$ 关上 的 [门]$_F$
shi4 dian4yuan2 guan1shang4 de0 **men2**
COP salesman close DE **door**
"It was [the salesman]$_F$ who closed [the **door**]$_F$."

Following the analysis of Büring (2003) (see also Constant, 2014; Riester, 2018), based on (4b) and (4c), (4a) sets up an implied question 谁关上了什么? "Who closed what?," which can be divided into two sub-questions 顾客关上了什么? "What did the customer close?' and 店员关上了什么? "What did the salesman close?," which is answered by the second focus in each of the following sentences. This part updates the common ground. However, importantly for our purposes, alternatives are implied for both the subject and object, as shown by the multiple focus-marking. As well as the alternatives to what was closed (窗户, 门 "window, door"), there are alternatives to the subject in the implied question, or contrastive topics (顾客, 店员 "customer, salesman"). That is, the syntactic and prosodic cues mark contrastive focus on different words (the subject and object respectively). While this kind of construction, a cleft with prosodic prominence in the main clause, has received little attention in the experimental literature, it is well attested in corpus-based studies of naturally occurring speech in English (Prince, 1978; Delin and Oberlander, 1995; Lambrecht, 2001; Hedberg, 2013), and it is shown to be used in certain contexts in natural speech in Chinese (Hole, 2011).

3. THE EFFECTS OF FOCUS ON LANGUAGE PROCESSING

In this section, we review the literature showing the effects of focus on the processing of focus-marked words. Almost all of this work has been on English and other Germanic languages, so most studies discussed are necessarily on these languages. We start by briefly reviewing studies looking at the processing of focus in general, and then review research on the role of focus-marking in word activation in lexical decision tasks, the method employed in this study.

It has long been established that focused words enjoy a processing advantage over unfocused or defocused words. These earlier studies assume a QUD-focus definition of focus. In phoneme-monitoring experiments, phonemes in focused words or the words whose preceding intonation contour predicts a focus are recognized faster (Cutler, 1976; Cutler and Fodor, 1979; Akker and Cutler, 2003; Ip and Cutler, 2017). Focused words are also remembered better (Birch and Garnsey, 1995; Birch et al., 2000; Kember et al., 2016a). For example, when a word in an *it*-cleft was presented later in a memory task, participants were faster in confirming that they had previously seen the word than when it was not in focus (Singer, 1976; Birch and Garnsey, 1995; Birch et al., 2000). These experiments used written stimuli, so the primary cue to focus was syntactic. However, it has been shown that readers generate "implicit prosody" while reading (Fodor, 1998, 2002; Stolterfoht et al., 2007; Jun, 2010; Jun and Bishop, 2015). Here, it is most likely the implicit prosody would have the nuclear accent in the cleft head (see above). Recently, Kember et al. (2016a) used a similar memory task to look at the effect of focus in spoken sentences in Korean. They found that both prosodic and syntactic cues to focus enhanced memory for focused words, with syntactic and syntactic+prosodic cues most effective.

More recently, another line of studies, using eye-tracking, have given the first psycholinguistic evidence of focus as facilitating activation of alternatives, i.e., for the contrastive focus definition of focus. These studies showed that contrastive accenting biases listeners to look at contrastive referents that are available in their visual display, compared to non-contrastive accenting which shows no bias (Dahan et al., 2002; Weber et al., 2006; Ito and Speer, 2008; Watson et al., 2008; Dennison, 2010; Kurumada et al., 2014). To our knowledge, though, the only two types of focus-marker investigated in these studies are contrastive accenting and focus particles (Kim, 2012; Kim et al., 2015). Another line of work has looked at whether different types of focus-marking facilitate memory for foci and their alternatives in discourse contexts (Fraundorf et al., 2010, 2013; Spalek et al., 2014). Fraundorf et al. (2010, 2013) found contrastive accents and font emphasis respectively facilitate memory for foci and correct rejection of alternatives, while Spalek et al. (2014) found focus particles facilitate memory for mentioned alternatives, but not focused words themselves, on accented words in discourse contexts.

Most importantly for our purposes, there have been studies looking at the role of focus-marking in word activation. Since the 1970s (Swinney et al., 1979), the activation of words given different linguistic primes has been investigated using cross-modal lexical decision tasks. These studies have shown that single words prime themselves (identity prime) and their semantic associates; but while identity priming is consistent in sentence contexts, semantic associative priming is not (Norris et al., 2006). For example, Norris et al. (2006) showed that when participants heard an auditory prime *seat*, they were quicker to recognize an identical printed target *seat* was a word (compared to an unrelated control target *river*); likewise they were quicker to respond to a semantically associated target *chair*. However, when the prime word was in a sentence, e.g., *He gave up the seat for me out of some form of courtesy*, participants were still faster to respond to the identical target *seat*, but not the semantic associate *chair*. Norris et al. (2006) then tested a number of variables that could affect semantic associative priming in sentence contexts. They found that this priming was only significant when the sentence was truncated immediately after the prime word, or when there was a contrastive accent in the sentence, whether or not this was on the prime word. They speculated that the latter result may be because the accent caused the listeners to attend to the sentence as a whole more carefully. They also suggest that the contextual relevance of the target to the meaning of the prime in the sentence may affect priming, although they do not directly link this to focus.

Braun and Tagliapietra (2010) provided a key insight into a possible reason for these results: whether the word is contrastively focused. According to alternative semantics theory, contrastive focus-marking should imply alternatives to the focused word. Hence, in a lexical decision task, alternatives should be activated when the prime in a spoken sentence is contrastively accented, but not when the prime is not. Braun and Tagliapietra (2010) compared semantic priming of the sentence-final object word in sentences with one of two intonation patterns in Dutch: contrastive, with contrastive accents on both the first and last content word in a sentence [e.g., (5a) and (5c)]; and neutral, with non-contrastive accents on these words [e.g., (5b) and (5d)]:

(5) a. In **Florida** he photographed a **flamingo** (**Contrastive - related prime**)
b. In Florida he photographed a flamingo (**Neutral - related prime**)
c. In **Florida** he photographed a **celebrity** (**Contrastive - control prime**)
d. In Florida he photographed a celebrity (**Neutral - control prime**).

In their first experiment, after hearing the prime sentence, participants saw a target(e.g., *pelican*). The object in the prime sentence was either related to, and was a contextual alternative, to *pelican*, e.g., *flamingo* in (5a) and (5b); or it was unrelated, e.g., *celebrity* in (5c) and (5d). Participants were quicker to decide *pelican* was a real word after hearing the related prime *flamingo* compared to the unrelated control prime *celebrity* when the sentence-final object was contrastively accented (alternative priming). However, there was no time advantage when the sentence-final object (*flamingo* or *celebrity*) was not contrastively accented. Their second experiment examined the priming of non-contrastive associates (e.g., *pink*) that were not plausible replacements for *flamingo*. They found that non-contrastive associates were weakly primed regardless of the prosody. The priming of contrastive and non-contrastive associates was not directly compared in the two experiments. However, it seems fair to say that alternatives were primed more than non-contrastive associates when the prime was contrastively accented.

Husband and Ferreira (2016) also looked at semantic priming in sentences with either contrastive or neutral accents, finding a somewhat different pattern of results to Braun and Tagliapietra (2010). In their study, the prime word was a sentence-medial object or adjective, e.g.,:

(6) a. The museum thrilled the **sculptor** when they called about his work (**Contrastive**)
b. The museum thrilled the sculptor when they called about his work (**Neutral**).

After hearing the sentence, participants saw a target which was either a contextual alternative (e.g., *painter*) or a non-contrastive associate (e.g., *statue*). Husband and Ferreira (2016) were interested in the time course of activation of the prime, or the stimulus onset asynchronies (SOA) from the prime word. In their first experiment, the SOA was 0 ms. This was similar to Braun and Tagliapietra (2010), however, as the prime word was non-final, this was while the sentence was still playing. Husband and Ferreira (2016) found all semantic associates were primed except for non-contrastive associates in the neutral prosody, as non-contrastive associates were less related to the semantic context and had less time to be activated. In their second experiment, the SOA was set at 750 ms. When the prime word was contrastively accented, the non-contrastive associates were responded to at the same speed as unrelated items while alternatives were faster, showing only the alternatives were primed. When the prime word had a neutral accent, both contrastive and non-contrastive associates were faster than the controls. Husband and Ferreira's (2016) explanation for the mechanism behind this was different to Braun and Tagliapietra (2010). They claim this shows all semantically related words are initially activated, but contrastive accenting prompts a selection mechanism whereby non-contrastive associates are rapidly deactivated, while contextual alternatives remain activated as they are likely to be relevant for interpretation. However, it should be noted that there were a number of other differences between the studies, including the details of how the contrastive/neutral accenting conditions were manipulated, and the time course of when the target was presented.

Braun and Tagliapietra (2010) and Husband and Ferreira (2016) only looked at priming of alternatives cued by contrastive accenting. A series of psycholinguistic experiments conducted by Gotzner and colleagues (Gotzner et al., 2016; Gotzner, 2017) explored the activation and processing of alternatives where the focus prime was marked by focus particles *only* and *even* in German. While contrastive accenting indicates the presence of relevant alternatives to the focus, focus particles add further semantic restrictions on the interpretation of those alternatives, e.g., *only* excludes the possible alternatives (similarly to the claimed effect of clefts discussed in section 2). Using both probe recognition tasks, and lexical decision tasks, they found focus particles slowed the recognition of mentioned alternatives, and the rejection of unmentioned alternatives. They attribute the result to an interference effect of the focus particle, due to increased competition between members of the alternative set.

Bringing together these studies, there is considerable evidence that focus, marked by contrastive accents, facilitates activation of alternatives to the focused word, in and out of a discourse context. However, there are a number of important gaps in our present knowledge of this process. Firstly, if focus-marking is the underlying mechanism, we should also expect identity priming to be strengthened for focus-marked words in sentence contexts, compared to non-focus-marked words. Focus-marked words should be more activated by either focus definition: they are part of the alternative set (Rooth, 1992), and they are important information as the QUD-focus (see the early findings on focus in phoneme-monitoring and memory tasks). While this has been shown in phoneme-monitoring and memory tasks (for contrastive accenting and clefting), it has not been looked in previous research using lexical decision tasks, to our knowledge. Secondly, the evidence is mixed as to whether focus affects priming in the absence of contrastive accenting. Syntactic focus-marking, i.e., clefting, was looked at in the earlier memory experiments (Birch and Garnsey, 1995; Birch et al., 2000), but not in the alternative priming studies (Braun and Tagliapietra, 2010; Husband and Ferreira, 2016). Focus particles, in addition to contrastive accenting, have been shown to slow processing in general, rather than further prime alternatives. Therefore, it is not yet established if other focus-marking mechanisms (e.g., clefts) facilitate priming of focused words and their alternatives in the absence of contrastive accents. If the underlying mechanism is focus-marking, this should be the case; however, it is not fully clear it is, and there are indications focus-marking apart from contrastive accenting can slow processing.

Thirdly, the priming effects of focus have only been looked at in a handful of closely related languages, i.e., English, Dutch, and German. It is therefore cross-linguistically important and interesting to see whether they can also be found in other language families, in this case, Mandarin Chinese. Very little research has been carried out in Mandarin Chinese on the processing advantage of focus. As discussed above, like English, in Chinese prosodic prominence is a primary marker of focus, although marked with phrasal prominence, rather than pitch accents. Therefore, we might expect these languages to be similar. In a phoneme-monitoring task in Chinese, Ip and Cutler (2017) showed that target phonemes in words were responded to faster when the preceding prosody predicted focus, in line with the findings in Germanic above (e.g., Cutler, 1976). More closely, our recent experiment (Yan et al., 2019) tested the role of contrastive prominence in priming focused words, contrastive alternatives and non-contrastive associates of subject nouns in canonical order sentences in Mandarin Chinese. The study followed a very similar design to the present one, except that the two sentence types compared were canonical word order sentences with contrastive prosodic prominence on the object (canonO in this study) or the subject (canonS, not included in this study). It was found that focused words and contrastive alternatives were recognized faster when the subject carried contrastive prominence (canonS) than when it did not (canonO). Non-contrastive associates were not primed in either of the conditions. However, we did not test the role of syntactic cues to focus.

In this paper, we report on a cross-modal lexical priming study testing the role of prosodic and syntactic focus-marking in facilitating priming of words and their alternatives in Mandarin Chinese. Note that for this study, focus-marking means contrastive focus: all of the focus-marking conditions compared have been shown to mark contrastive focus on the subject, but not necessarily QUD-focus (although some also mark QUD-focus), see section 2. The prime word was always the subject noun, with the target presented after the end of the sentence. We were interested in the priming effects after a longer course of processing, rather than immediate processing, as this is when effects of focus should be stronger (as per Husband and Ferreira, 2016). Braun and Tagliapietra (2010) and Husband and Ferreira (2016) looked at sentence-final (objects) or sentence-medial elements, so to our knowledge no studies in this area have yet tested subject nouns using cross-modal lexical priming paradigms. Subject nouns are interesting to look at, as previous work has shown that positional cues to focus affect processing ease (e.g., Repp and Drenhaus, 2015).

4. THE EXPERIMENT

4.1. Research Questions

This experiment addressed the following research questions:

1. Is prosodic or syntactic F(ocus)-marking necessary for subject nouns to prime *themselves*? If not, do they strengthen the priming?
2. Is prosodic or syntactic F-marking necessary for subject nouns to prime their *contrastive alternatives*? If not, do they strengthen the priming?
3. Is prosodic or syntactic F-marking necessary for subject nouns to prime their *non-contrastive associates*? If not, do they strengthen the priming?

4.2. Methods

4.2.1. Participants

Ninety-nine (79 females and 20 males) native Mandarin Chinese speakers (mean age = 20.77, SD = 1.92, age range = 18–26) were recruited from students at Henan Polytechnic University in China. 80 were from Henan province and 19 were from other Mandarin speaking provinces. They reported that they had received English education, but they did not speak other languages at home and were not fluent in any other languages. They had not lived outside China for more than 6 months. They received supermarket vouchers in recognition of their participation. None of them reported any hearing or reading difficulties.

4.2.2. Materials and Design

Sixty critical sentences were constructed containing a prime word as the subject noun (see full list in the **Supplementary Material**). All sentences described a simple, plausible event in the past tense, using commonly occurring nouns and verbs. As much as possible, the event described by the verb and the object was not semantically related to the subject, so there were no potential semantic priming relationships within the sentence. Most of the sentences were subject-verb-object (SVO) sentences; six were subject-verb-preposition-object. They all had seven syllables in the canonical order version.

For each sentence, three sentence type versions were created, involving different focus-marking on the subject noun (see examples in **Table 1**): *no F-marking*, i.e., canonical word order with nuclear prominence on the object (canonO); *syntactic F-marking*, i.e., subject cleft with nuclear prominence on the object

TABLE 1 | Sentence types, with F-marking, and target types used in the Chinese experiment (bold shows nuclear prominence; the information on F-marking refers only to the subject noun in each case).

Sentence types	Examples
canonO (no F-marking)	顾客关上了**窗户** "The customer closed the **window**."
ScleftO (syntactic F-marking)	是顾客关上的**窗户** "It was the customer who closed the **window**."
ScleftS (prosodic+syntactic F-marking)	是**顾客**关上的窗户 "It was the **customer** who closed the window."
Target types	**Examples**
Identical	顾客 "customer"
Contrastive	店主 "shop owner"
Non-contrastive	产品 "product"
Control	陆地 "land"

(ScleftO); and *prosodic+syntactic F-marking*, i.e., subject cleft with nuclear prominence on the subject. For each sentence, a quadruplet of four target types was constructed (identical item, contrastive alternative, non-contrastive associate, unrelated control). The contrastive alternatives could replace the subject nouns in the sentence. The non-contrastive associates were related to the subject nouns, but could not replace them in the sentence. The unrelated controls were not related to the subject nouns. All target words were not related to the objects and verbs to avoid being primed by them. Three sentence types and four target types resulted in twelve experimental conditions. 60 sentences were used to make 180 experimental sentences (60 sentences * 3 sentence types). Each sentence was paired with four target types, which gave a total of 720 experimental stimuli. Twelve lists of 60 experimental stimuli were constructed in a Latin square design. Each participant saw only one list.

There were several further steps involved in preparing the experimental stimuli, which are described below. First, we describe a survey carried out to create semantic relatedness norms needed to control for semantic relatedness between target types. Second, for similar reasons, we controlled for word frequency between words across target types. Then we describe the recording and acoustic analyses of the experimental stimuli. Finally, we describe the construction of other items (fillers).

4.2.2.1. Relatedness scores

The semantic relatedness between the non-identical targets and the subject nouns was tested, to be able to control for this in the analysis. Since there were no published association norms for Mandarin, the relatedness scores were collected from an online questionnaire constructed in Qualtrics (2017).

Seventy-five common disyllabic nouns were extracted from a Chinese word frequency corpus SUBTLEX-CH (Cai and Brysbaert, 2010). Seventy-five short sentences were constructed with the nouns as the subject. Then three other words were selected from the corpus for each sentence: contrastive alternative, non-contrastive associate and unrelated control, relative to the subject noun. These words were not related to, and could not replace, any other word in the sentence. Therefore, there were 75 quadruplets, each resulting in three pairs of ratings: subject noun vs. contrastive alternative, subject noun vs. non-contrastive associate, subject noun vs. unrelated control. Sixty-seven native Mandarin speakers from Henan Polytechnic University completed the online questionnaire. Each participant saw only one of the three pairs. They were asked to rate the relationship between two words from 1 "*not related at all*" to 7 "*highly related*" in the presence of a context sentence (e.g., *how related are "customer" and "salesman" in the sentence "The customer closed the window"*). Yan (2017) showed that context affects the relatedness scores. The participants who took part in the online questionnaire did not participate in the lexical decision task.

Following the survey, 60 sentences were chosen in order to have similar relatedness scores between the subject noun and both of the two types of associates, and also for the subject noun and the unrelated control to be as unrelated as possible. The mean relatedness score was 4.83 (SD = 1.88) for prime-contrastive (e.g., *customer-shop owner*), 5.05 (SD = 1.89) for prime-non-contrastive (e.g., *customer-product*), and 1.77 (SD = 1.34) for prime-unrelated (e.g., *customer-land*). Relatedness scores as the ordinal dependent variable and the relationship between two words in a pair as the independent variable were fitted into a cumulative link mixed model using the `ordinal` package in (R Core Team, 2017; Christensen, 2019). The results showed no significant differences between the prime-contrastive pair and the prime-non-contrastive pair ($z = -1.586, p = 0.26$). However, significant differences were found between the prime-control pair and the prime-contrastive pair ($z = 21.17, p < 0.001$) and between the prime-control pair and the prime-non-contrastive pair ($z = 22.00, p < 0.001$).

4.2.2.2. Frequency

The log frequency of each target word was collected from SUBTLEX-CH (Cai and Brysbaert, 2010). The mean log frequency of the chosen items was 3.085 (SD = 0.44) for subject nouns, 2.916 (SD = 0.43) for contrastive alternatives, 2.790 (SD = 0.43) for non-contrastive associates and 2.917 (SD = 0.44) for unrelated controls. The log frequency of each word was fitted into an ANOVA, and the *post-hoc* Tukey test showed a significant difference between subject nouns and non-contrastive associates (e.g., *product-shop owner*) [$t_{(236)}$=3.70, $p = 0.002$], though the frequencies between word types were controlled to be closely matched. No significant differences were found between the other groups (all p-values > 0.1).

4.2.2.3. Recording and acoustic analysis

The sentences were recorded directly to hard drive using Praat (Boersma and Weenink, 2018) by a trained female native Mandarin speaker (first author) in a soundproof room at Victoria

University of Wellington through a USB-based microphone (see **Figure 1** above for examples of canonS and canonO, and **Figure 2** for ScleftS and ScleftO). All sentences were checked impressionistically by two native Mandarin speakers for the location of prosodic prominence.

The acoustic measurements (duration, mean F0, max F0, min F0 and mean intensity) of words were obtained using ProsodyPro (Xu, 2013). As focus is marked through pitch range expansion in Chinese, F0 range was also calculated being the difference between max F0 and min F0. The measurements (duration, mean F0, F0 range, and mean intensity) were fitted as the dependent variable in separate linear mixed effects models, using the R package `lme4` (Bates et al., 2015). The fixed effects initially included sentence type (canonO, ScleftO, ScleftS) and word position (subject, object) as well as the interaction between the two. Tone combination was also included, as tone affects syllable duration and F0 (e.g., Long, 1985). Word was the random effect. Each model was reduced to remove non-significant factors (see further section 4.2.4). The fitted values are provided in **Table 2**. The ANOVA tables of the final models for each measurement are provided in **Table 3**. Tone combination was a significant factor for duration, mean F0 and F0 range. All four models showed a significant interaction between sentence type and word position. In general, as **Table 2** shows, in subject-stressed sentences, the subject was more prominent than the object, whereas in object-stressed sentences, the object was more prominent than the subject. Planned comparisons, which were run using the `emmeans` function in the `lsmeans` package (Lenth, 2016), showed that, within the same sentence type, prosodically focused subjects or objects were more prominent than unfocused subjects or objects in terms of all four parameters (all p-values < 0.05). Across sentence types, subject words in the subject-stressed sentence type (ScleftS) had longer duration, higher F0, larger F0 range, and higher intensity than those in the object-stressed sentence type (canonO and ScleftO) (all p-values $<$ 0.05). Moreover, object words in ScleftS were less prominent than those in canonO and ScleftO (all p-values < 0.05). The aforementioned differences confirm that the materials were created as intended.

Stimulus onset asynchrony (SOA), the duration between the offset of the prime word and the onset of the visual target, was shown to influence the priming of target words in Husband and Ferreira (2016), compared to no SOA (0 ms). In order to keep the SOA constant, a variable duration of silence (0 ms to 607 ms) was added to the end of each sound file, so that the SOA was always 1,500 ms.

TABLE 2 | Fitted mean values of duration (ms), F0 (Hz), F0 range (Hz), and intensity (dB) of subject and object nouns in Chinese critical stimuli.

Sentence condition	Word position	Duration	F0	F0 range	Intensity
canonO	Subject	566	216	81	70
	Object	740	288	243	75
ScleftO	Subject	535	210	72	70
	Object	732	283	243	75
ScleftS	Subject	680	336	264	79
	Object	585	180	85	64

4.2.2.4. Other items

A further 150 filler sentences with word and non-word targets were constructed, which lead to a total of 210 trials per list (60 test items + 150 fillers). As the experiment task is to decide whether two characters make up a real word or not in Mandarin Chinese, we included non-words to avoid response bias. Among these filler targets, 105 were non-words and 45 were words to counterbalance yes/no responses across the whole experiment. Sixty of the filler sentences had the same sentence types (canonO, ScleftO, ScleftS) with non-words as target words. Among the non-words, ten were phonologically related to one of the words in the sentence to encourage different types of priming. Another 90 filler sentences with different sentence structures (SV, SVAdv etc.) were also constructed, including 45 sentences with words and 45 with non-words as visual targets. Ten words and 10 non-words were phonologically related to one of the words in the sentence. Non-words were selected from the lexical decision data from Cai and Brysbaert (2010) with 100% non-word accuracy. The non-words consist of two real characters which do not make up a real word together. Six practice sentences which had three word and three non-word visual targets were also prepared. Furthermore, 12 comprehension questions asking the content of a previous filler were included to encourage participants to pay attention to the sentences.

4.2.3. Procedure

The experiment was administered using Opensesame v. 3.1 (Mathôt et al., 2012), and was run in a quiet computer room at

TABLE 3 | The ANOVA tables for duration, F0, F0 range, and intensity analysis.

	Chisq	Df	*P*
Duration: [model:duration ∼ SentenceType*wordPosition+ ToneCombination+(1|word)]			
SentenceType	15.87	2	<0.001
WordPosition	103.56	1	<0.001
ToneCombination	40.01	19	0.003
SentenceType:wordPosition	738.84	2	<0.001
F0: [model:F0 ∼ SentenceType*wordPosition+ToneCombination+(1|word)]			
SentenceType	18.02	2	<0.001
WordPosition	0.24	1	0.621
ToneCombination	139.99	19	<0.001
SentenceType:wordPosition	1781.49	2	<0.001
F0 range: [model:F0 range∼ SentenceType*wordPosition+ ToneCombination+(1|word)]			
SentenceType	6.64	2	0.036
WordPosition	45.57	1	<0.001
ToneCombination	70	19	<0.001
SentenceType:wordPosition	666.54	2	<0.001
Intensity: [model:intensity ∼SentenceType*wordPosition+(1|word)]			
SentenceType	32.70	2	<0.001
WordPosition	6.51	1	0.011
SentenceType:wordPosition	2950.44	2	<0.001

Henan Polytechnic University. The entire session was conducted in Chinese. Participants were seated in front of a computer screen with a closed-ear headphone. At the start of the experiment, participants received written instructions on the computer screen, and the instructions were also repeated orally by the experimenter (first author) after the participants had read them. In the practice phase, participants first heard a sentence, and while the sentence was being played, participants concentrated on a fixation dot in the middle of the screen. Then they saw two characters, and had to decide whether these two characters made up a real word or not by pressing "m" key [labeled as 是("yes")] for *yes* response and "z" key [labeled as 否("no")] for *no* response using their dominant hand as fast as they could. In the practice phase, participants received feedback on their responses (if their answer was wrong) and reaction times (RTs) (if their response time exceeded 1,000 ms).

The procedure of the main experiment was similar to the practice phase, but no feedback was provided. The main experiment moved to the next trial automatically if no key was pressed within 3 s. The stimuli were divided into four blocks with a 10 s compulsory break, or longer if participants wanted, between two blocks. The stimuli within a block were randomized as well as the order of blocks. Twelve filler trials were followed by the twelve comprehension questions which appeared randomly and evenly across the four blocks. The comprehension questions required "x" or "n" key press to adjust to the comprehension questions being a different task (from lexical decision) and therefore avoid mistakes. There was always a filler trial following the comprehension question. The entire experiment lasted approximately 15 minutes. Demographic information such as sex, age, hometown, and English proficiency was collected using a paper form at the end of the experiment.

4.2.4. Analysis Method

Both accuracy and response times (RTs) were measured. The accuracy measure enabled us to look at whether different focus conditions and target types had any influence on the difficulty of the lexical decision. RTs of lexical decisions reflected the activation of the visual target word by the auditory prime sentence. As priming was of central interest in the study, we primarily looked at the RTs. The comparison of RTs to the related words (identical, contrastive, non-contrastive) to unrelated baseline controls shows whether the related words were primed or not.

Mixed effects regression models were built to test how the accuracy and RTs were affected by a number of factors, using the R package `lme4` (Bates et al., 2015). For the accuracy analysis, response choice was the dependent variable in generalized linear mixed effects models (family: binomial) and for the RT analysis, reaction times were the dependent variable in linear mixed effects regression.

The fixed effects of the initial model included key experimental predictors and item factors. The key experimental predictors were sentence type (canonO, ScleftO, ScleftS) and target type (identical, contrastive, non-contrastive, and control). Backward difference coding was used to better represent the internal structure of sentence types, resulting in two variables: syntax (canonO vs. ScleftO) and prosody (ScleftO vs. ScleftS). The item factors included the log frequency of target words, the centered position of the trial in the experiment, and the transformed RTs of the previous trial, as these factors have been previously shown to influence RTs (e.g., Braun and Tagliapietra, 2010; Gotzner, 2017). Similarly, whether the previous target was a word and whether the previous response was correct were included as they can have spillover effects on the subsequent trial. Silence duration was also included as a predictor in the model, as this was variable between stimuli[2].

In addition to the fixed effects, the random effects, motivated by the literature and justified by the data, included intercepts for participants and target words, random slopes for trial (position in the experiment) by participants and by items and random slopes for the interactions between the key experimental factors by participants and by items. If the initial model did not converge, the model was simplified by reducing random structures, i.e. taking out the slopes that had the lowest variance scores until the model converged. When the model converged, the `step` function in the `lmerTest` package (Kuznetsova et al., 2017) was used to eliminate non-significant fixed and random effects. Only the factors that significantly contributed to the model fit were kept in the final model.

4.3. Results

A total of 20,790 responses were recorded, 210 from each of 99 participants. The overall accuracy is 91.5% for responses and 98.2% for comprehension questions. Data from three participants was excluded for low accuracy on target word responses, and one further as the "yes" button was not pressed with the dominant hand. The remaining 5,700 critical trials from 95 participants were used for accuracy analysis. A further 123 trials with incorrect responses (2.2%) were excluded, leaving 5,577 for the response time analysis. Further, data points of residuals whose standard deviations were larger than 2.5 were eliminated. The RTs were inverse transformed, which was the best transformation (that had the highest correlation in a quantile-quantile plot of the distribution), compared with no transformation, log transformation and inverse square root transformation. The transformed RTs were then multiplied by 10,000 in order to make the estimates and SD more readable.

4.3.1. Accuracy

The overall accuracy on the experimental trials was 97.8%. Mixed effect logistic regression models were built to test the factors affecting accuracy, following the process detailed above in section 4.2.4. The final model did not include sentence type, and had a random effect of Participant only. The fitted accuracy

[2]Note that silence duration was correlated with sentence type, as canonO and ScleftO sentences were longer than ScleftS sentences because of the stress on the object noun. We therefore initially regressed silence duration against sentence type [$F_{(2)} = 2710$, $p < 0.001$]. The residuals (*Residual Silence*) (the difference from the mean silence duration for each sentence type) were used as a predictor. But one reviewer pointed out the issues with residualization as a way to deal with colinearity (see Wurm and Fisicaro, 2014). We therefore used the unresidualized variable in the model. It should be noted that neither residualized nor unresidualized variable significantly improved the model fit ($p > 0.1$), so we did not keep it in the final model.

TABLE 4 | Fixed effects of mixed effects model with accuracy or transformed reaction times as the dependent variable.

	Chisq	Df	*P*
Accuracy: [model:correct ∼ TargetType+log frequency+centerd trial +(1|Participant)]			
TargetType	13.34	3	<0.001
log frequency	11.12	1	<0.001
centered trial	10.25	1	0.001
RTs: [model:transformed RTs ∼ Sentence condition*TargetType +log frequency+centerd trial+PreCorrectness+PreRT+PreWordness +(1|Participant)+(1|Item)]			
Sentence condition	18.11	2	< 0.001
TargetType	49.38	3	< 0.001
log frequency	36.21	1	< 0.001
centerd trial	90.99	1	< 0.001
PreCorrectness	9.05	1	0.003
PreRT	181.82	1	< 0.001
PreWordness	23.96	1	< 0.001
Sentence condition:TargetType	14.96	6	0.021

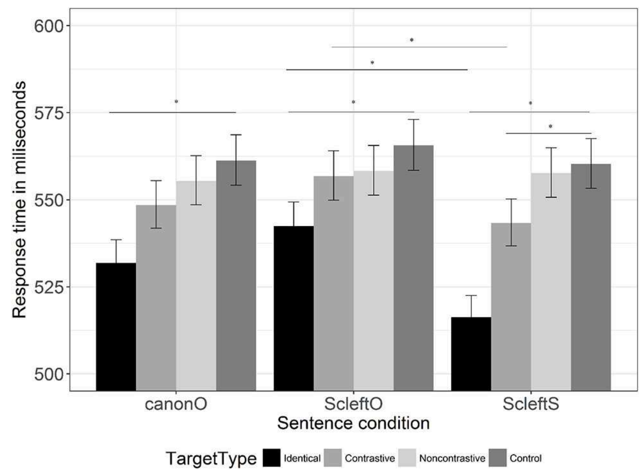


FIGURE 3 | Back-transformed fitted RTs in ms to four target types in canonO, ScleftO, and ScleftS conditions. Error bars show standard error of the means. Stars (*) show significant comparisons ($p < 0.05$).

was 99.1% for identical, 98.9% for contrastive, 98.6% for non-contrastive, and 97.8% for controls. The ANOVA table of the final model showing the significance of the fixed effects is in **Table 4**. Participants were more accurate for more frequent targets (β = 0.73, SD = 0.22) and later in the experiment (β = 0.01, SD < 0.01). In order to test which target types differed from each other, we conducted planned comparisons using the `glht` function in the `multcomp` package (Hothorn et al., 2008). Identical and contrastive items received higher accuracy rates than control items (identical: $z = 3.1$, $p = 0.01$; contrastive: $z = 2.8$, $p = 0.03$), but no significant differences were found between other target types (identical vs. contrastive; identical vs. non-contrastive; contrastive vs. non-contrastive; all p-values > 0.1).

4.3.2. Reaction Times

Mixed effect linear regression models were built to test the factors affecting RTs, following the process detailed in section 4.2.4. The ANOVA table showing the significance of variables in the final model is given in **Table 4**. The final model had random effects for Participant and Item. Participants became faster over the course of the experiment (centered trial: $\beta = 0.006$, SD = 0.001). Since the dependent variable is an inverse transform of RT, negative coefficient estimates represent slower responses, and positive coefficient estimates represent faster responses. Words of higher frequency were recognized faster (log frequency: β = 0.82, SD = 0.14). Participants responded more quickly when the previous response was correct (PreCorrectness: β = 0.57, SD = 0.19); and when the previous trial was a word (PreWordness: β = 0.39, SD = 0.08). Participants responded more slowly when the transformed RT to the previous trial was slow (PreRT: β = –0.003, SD = <0.001). None of the other factors included in the initial model were significant (see section 4.2.4), thus we will not discuss them.

The final model showed main effects of sentence condition and target type, as well as their interaction (see **Table 4**). The fitted RTs are shown in **Figure 3**. As expected, identical words were recognized the fastest at 530 ms, then contrastive alternatives (549.5 ms), and then the other two target types (non-contrastive: 557.1 ms; control: 562.4 ms). For sentence condition, ScleftS was the fastest (543.1 ms), followed by canonO (548.3 ms) and ScleftO (555.6 ms). In order to find out how different target types were affected by sentence condition, we conducted planned comparisons on the interaction using the `glht` function in the `multcomp` package. In order to run the comparison, the model was rerun with the interaction between sentence condition and target type as a single factor.

To investigate the first research question: whether prosodic or syntactic F-marking is necessary for subject nouns to prime *themselves*, we conducted planned comparisons between identical items (subject nouns) and unrelated controls in the no F-marking condition (canonO), the syntactic F-marking condition (ScleftO) and the prosodic+syntactic F-marking condition (ScleftS). Identical items showed facilitation over unrelated controls in all sentence conditions (canonO: $z = 4.37$, $p < 0.001$; ScleftO: $z = 3.37$, $p = 0.003$; ScleftS: $z = 6.79$, $p < 0.001$). This indicates that F-marking is not necessary for subject nouns to prime themselves, as subjects nouns were recognized faster than unrelated controls in the no F-marking condition.

To investigate the second research question: whether prosodic or syntactic F-marking is necessary for subject nouns to prime their *contrastive alternatives*, we conducted planned comparisons between contrastive alternatives and unrelated controls in the three focus conditions. Contrastive alternatives were facilitated over unrelated controls in the ScleftS condition (ScleftS: $z = 2.5$, $p = 0.043$), but not in the other two conditions (canonO: $z = 1.86$, $p = 0.135$; ScleftO: $z = 1.25$, $p = 0.358$). This shows that prosodic F-marking is necessary for subject nouns to prime contrastive alternatives, as contrastive alternatives were only recognized faster than unrelated controls in the ScleftS condition.

To investigate the third research question: whether prosodic or syntactic F-marking is necessary for subject nouns to prime their *non-contrastive associates*, we also conducted planned comparisons between non-contrastive associates and unrelated controls in the three focus conditions. None of the comparisons were significant (all p-values > 0.1). This showed that the non-contrastive associates were not primed in any of the conditions.

We also did planned comparisons between the syntactic F-marking condition and the no F-marking and the prosodic+syntactic F-marking condition (canonO vs. ScleftO; ScleftS vs. ScleftO) for all four target types. The results showed that only identical items and contrastive alternatives were facilitated in the prosodic+syntactic condition compared to the syntactic condition, which showed that prosodic F-marking strengthened the priming of identical items ($z = 4.96$, $p < 0.001$), and that prosodic F-marking primed contrastive alternatives ($z = 2.4$, $p = 0.047$). All the other comparisons were not significant (all p-values > 0.1).

Table 5 summarizes the comparisons laid out above that were relevant and important to answer the research questions, e.g., identical words were primed (relative to unrelated controls) in all sentence conditions, and the priming was strengthened in the ScleftS condition. Contrastive alternatives were primed in the ScleftS condition, but not in canonO and ScletO conditions. Non-contrastive associates were not facilitated over unrelated controls in all sentence conditions.

We also ran an additional analysis to test the effects of the relatedness of the prime word to the visual target, using the relatedness scores from our questionnaire (see section 4.2.2). This analysis excluded trials with identical targets, as this would be between a prime word and itself. An ANOVA model comparison showed that relatedness did not significantly improve the model fit[$\chi^2_{(1)}$=2.04, $p = 0.15$].

5. GENERAL DISCUSSION

We reported a cross-modal lexical decision experiment, looking at the priming of different kinds of targets in Mandarin Chinese. Primes were subject nouns in spoken sentences. Targets were presented after the sentences, with a fixed SOA of 1,500 ms. The experiment looked at four target types: identical items, contrastive alternatives, non-contrastive associates, and unrelated controls; and three sentence types: no focus-marking (canonO, canonical order with nuclear stress on the object), syntactic focus-marking (ScleftO, subject cleft with nuclear stress on the object), or prosodic+syntactic focus-marking (ScleftS, subject cleft with nuclear stress on the subject). The study addressed three main questions (see section 4.1): whether prosodic or syntactic focus-marking is necessary for subject nouns to prime *themselves*, *their contrastive alternatives*, and *non-contrastive associates*.

TABLE 5 | Comparisons of related words (identical, contrastive, non-contrastive) and unrelated controls in all three sentence conditions (canonO, ScleftO, ScleftS).

Target types	canonO (no F-marking)	ScleftO (syntactic F-marking)	ScleftS (syntactic + prosodic F-marking)
Identical vs. control	*	*	*
Contrastive vs. control	NS	NS	*
Non-contrastive vs. control	NS	NS	NS

Star (*) show significant comparisons (p < 0.05).

In relation to the first research question, subject nouns in spoken sentences prime themselves in Mandarin Chinese (identity priming). Identical words were responded to significantly faster than unrelated controls in all conditions. Further, the priming effect was strengthened by prosodic focus-marking (see **Figure 3**). This is consistent with the effect of prosodic focus-marking, in the absence of syntactic focus-marking (i.e., canonS vs. canonO), on identical priming reported in Yan et al. (2019). There it was also found that identical items were responded to faster when they were prosodically prominent (prosodic focus-marking). This shows that focus-marking is not necessary for subject nouns to prime themselves; however prosodic focus-marking, but not syntactic focus-marking, strengthens the priming. The general result that identity priming is found in all focus conditions is consistent with Norris et al. (2006) for English, and validates the effectiveness of the methodology in Chinese. Together with our results reported in Yan et al. (2019), this shows for the first time that identity priming is strengthened by prosodic focus-marking in Chinese.

In relation to the second research question, contrastive alternatives were recognized significantly faster than unrelated controls in the prosodic+syntactic focus-marking condition, but not in the no focus-marking and syntactic focus-marking conditions. Therefore, prosodic focus-marking is necessary for subject nouns to prime their contrastive alternatives, which is consistent with the findings in Yan et al. (2019) in the absence of syntactic focus-marking in Mandarin. This is also consistent with what Braun and Tagliapietra (2010) found for Dutch, but is different to what Husband and Ferreira (2016) found for English, who found that contrastive alternatives were responded to faster than controls in both the neutral and contrastive accenting conditions in English (with an SOA of 750 ms, which is closest to our experiment). In our study, syntactic focus-marking did not play a similar role in the priming of contrastive alternatives, as contrastive alternatives were not recognized faster than unrelated controls when the subject nouns were marked with syntactic focus-marking.

In relation to the third research question, the RTs of non-contrastive associates were not significantly different from those of unrelated controls in any sentence condition, nor did the RTs for non-contrastive associates significantly differ by sentence condition, showing they were not primed. RTs for non-contrastive associates were, however, numerically faster than for controls across conditions. This result is again consistent with that found in Yan et al. (2019), which also showed no difference in RTs between non-contrastive associates and unrelated controls regardless of prosodic focus-marking (canonS vs. canonO). Concerning the role of prosodic focus-marking in priming non-contrastive associates, this result is largely consistent with Braun and Tagliapietra (2010), who found weak priming of non-contrastive associates regardless of sentence conditions in Dutch. Even though non-contrastive associates had only numerical facilitation with prosodic focus-marking in our study and Husband and Ferreira, our finding is different from Husband and Ferreira (2016), who only found priming of non-contrastive associates in later processing (SOA 750 ms) in their neutral accent condition, but no priming of non-contrastive associates with prosodic focus-marking.

These results therefore provide cross-linguistic psycholinguistic evidence for the role of prosodic focus-marking in lexical activation. They extend previous findings, using phoneme monitoring and memory tasks, that prosodic focus-marking increases attention to and activation of the focused word, by showing prosodic focus-marking strengthens identity priming. Further, together with the results in Yan et al. (2019), they show for the first time in a non-Germanic language, evidence for prosodic focus-marking as activating alternatives to the focused word, consistent with Rooth's (1992) theory, by showing that prosodic focus-marking primes contrastive alternatives to subject nouns in Chinese, in canonical order and cleft sentences. These findings are consistent with those for Dutch reported in Braun and Tagliapietra (2010), and related findings using eye-tracking and other findings in Germanic languages reported in section 3. The differences between contrastive and non-contrastive associates in the Chinese experiment and earlier studies show this is not a general semantic priming effect, but is rather consistent with the role of prosodic focus-marking in triggering an implication of alternatives.

Our results on the role of prosodic focus-marking in priming contrastive alternatives are consistent with those found for Dutch by Braun and Tagliapietra (2010), in that contrastive alternatives were only primed with contrastive prosody, but are different from those found for English by Husband and Ferreira (2016), who found priming of contrastive alternatives with neutral or contrastive accenting. There were some methodological differences between the earlier studies and ours, e.g., in how semantic relatedness between target types was controlled, and in relation to the timing of the presentation of the targets (SOA). These were presented immediately after the object prime in Braun and Tagliapietra (2010), with an SOA of both 0 ms and 750 ms in Husband and Ferreira (2016), and with an SOA of 1,500 ms in this study. The time course therefore does not seem to account for the difference in results for priming of contrastive alternatives, but rather suggests that contrastive alternatives remain activated for a long time course. This is consistent with the facilitation results in memory tasks reported in section 3. There were also differences between the studies in the prosodic realization of the "neutral" or "no prosodic marking" condition. In both Braun and Tagliapietra (2010) for Dutch and Husband and Ferreira (2016) for English, the prime word in their "neutral" accent condition was in fact still accented. In Dutch, this was an !H* accent at the end the Dutch "hat pattern", with steady, low or falling pitch through the object word; whereas in English, the (!)H* accent was a definite rise, although small. Thus, the former may have been less prosodically prominent than the latter. In the Chinese stimuli for canonO and ScleftO, the pitch range was relatively narrow for the subject, and much wider for the object. Therefore, we speculate that the Chinese and Dutch "neutral"/"no focus" stimuli were more similar in terms of prosodic realization.

On the other hand, our results in relation to non-contrastive associates were more similar to Husband and Ferreira (2016). Braun and Tagliapietra (2010) found weak priming of non-contrastive associates regardless of prosody at 0 ms SOA, and Husband and Ferreira (2016) found priming only with contrastive accenting at 0 ms, and only without contrastive accenting at 750 ms, while we found no priming, regardless of prosody, at 1,500 ms SOA. In this case, the time course of presentation does seem like the most likely reason for the differences in results. As discussed in section 3, general semantic priming is not consistent, and it may be shorted-lived (see e.g., Neely, 1977). Husband and Ferreira (2016) account for their results in terms of rapid decay of general (non-contrastive) semantic associates, which is expedited by contrastive accenting. Considering our SOA was even longer (1,500 ms), it may be that general semantic priming had decayed over this time course, regardless of prosody. It is also possible that these different findings for contrastive and non-contrastive associate priming stem from language-specific differences in processing, though there is no obvious reason for the particular differences between Dutch, English, and Chinese found.

In Chinese, syntactic focus-marking without prosodic prominence (ScleftO) seemed to slow recognition times in general; although the differences were not significant. One reason might be that ScleftO sentences usually require a context, such as (4a) in section 2, where the subject is presupposed. The relative unusualness out of context might have slowed responses. These findings resemble those in Gotzner (2017), who found the exclusive focus particle *only* also slowed listeners' response times. She argued that focus particles had interference effects caused by stronger competition among members of the alternative set. Similarly, here responses could be slowed by the difficulty of encoding the presuppositions required by the ScleftO structure. On the other hand, Gotzner showed that in memory, *only* had a processing advantage. It could be that more complex ways of marking focus have an immediate processing cost, but a later processing advantage. In future work it would be good to look at the effect of syntactic focus in memory tasks.

What does this mean for the relationship between different types of focus-marking and lexical activation, given that we have found that prosodic focus-marking, but not syntactic focus-marking, strengthens activation of focused words and is necessary for alternative priming in Chinese? If focus-marking, and not specifically prosodic focus-marking, is the underlying mechanism, it is surprising that syntactic focus-marking did not strengthen priming. Perhaps contrastive prosodic prominence is the underlying mechanism, and the focus effect observed in previous studies with written syntactic clefting may be triggered by the implicit contrastive prominence. Therefore, it is possible that the activation is rather related to prosodic prominence, which enhances the salience of the prominent word, and therefore its processing, rather than focus-marking. However, it is also possible that prosodic and syntactic focus-marking play different roles, at least in Germanic and Chinese. As discussed above, syntactic clefting (and other morphosyntactic markers like focus particles) carry additional implications, such as more complex presuppositions and exhaustivity. This may slow processing in the short term, but have memory advantages. A further reason for the finding could be because, in Chinese (and Germanic), prosody is the primary cue to focus, while syntax is secondary (and hence carries additional implications). If we were to look at languages where morphosyntactic markers were the primary cue to focus, and prosody secondary, we would expect to see strengthening of activation and priming of alternatives given those morphosyntactic markers. We need more studies to distinguish between these possible explanations, but these results suggest this is a fruitful area for future research.

This study aimed to shed light on the role of focus-marking in lexical activation, and particularly, in the linguistic cues which listeners use to activate alternatives in spoken sentences. This is an important part of understanding the processes by which listeners understand implicatures related to alternatives in speech. The results further strengthened earlier findings for the importance of prosodic prominence in strengthening activation of focused words and their alternatives, and importantly, provided cross-linguistic validation of this in Chinese. However, it revealed a complex picture of the cues which strengthen identity priming and trigger alternative priming, i.e., prosody but not necessarily syntax. We hope this will prompt more work on the linguistic cues to focus listeners attend to in speech, and their apparently highly contextual nature.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Victoria University of Wellington Human Ethics Committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Victoria University of Wellington Human Ethics Committee.

AUTHOR CONTRIBUTIONS

This study was originally conceived by SC. The experiment was mainly designed and built by MY with contributions by SC. Both authors contributed to the data collection and data analysis and the write up of the paper.

FUNDING

This work was supported by the Marsden Fund Council (grant number 15-VUW-015) from New Zealand Government funding, awarded to SC, and a Faculty of Humanities and Social Sciences Research Grant (grant number 218221), Victoria University of Wellington, awarded to MY.

ACKNOWLEDGMENTS

Thanks to Paul Warren for much useful discussion of the design and analysis of the study. Thanks to three reviewers, and the guest editors, for helpful comments on earlier versions of the paper. Thanks also to the audience at the ProsLang Workshop, held at Victoria University of Wellington in November 2018, for useful discussion of an earlier presentation of these findings. Thank you to Minxia Wang at Henan Polytechnic University for hosting the first author, and help in recruiting participants. Thanks to Lisa Woods for help in advising on the analysis of the data. Finally thank you to all participants for taking part.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01985/full#supplementary-material

The sound files for the eight examples shown in **Figures 1** and **2** are included as supplemental material, along with the full list of stimuli, relatedness scores, and word frequencies for the experiments.

REFERENCES

Akker, E., and Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Biling. Lang. Cogn.* 6, 81–96. doi: 10.1017/S1366728903001056

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Birch, S. L., Albrecht, J. E., and Myers, J. L. (2000). Syntactic focusing structures influence discourse processings. *Discourse Process.* 30, 285–304. doi: 10.1207/S15326950dp3003_4

Birch, S. L., and Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *J. Mem. Lang.* 34, 232–267. doi: 10.1006/jmla.1995.1011

Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer [Computer Program]*. Version 6.0.37. Available online at: http://www.praat.org.

Braun, B., and Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Lang. Cogn. Process.* 25, 1024–1043. doi: 10.1080/01690960903036836

Büring, D. (2003). On D-trees, beans and B-accents. *Linguist. Philos.* 26, 511–545. doi: 10.1023/A:102588770

Cai, Q., and Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE* 5:e10729. doi: 10.1371/journal.pone.0010729

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: a probabilistic perspective. *Language* 86, 1–42. doi: 10.1353/lan.0.0197

Chafe, W. (1976). "Givenness, contrastiveness, definiteness, subjects, topics, and point of view," in *Subject and Topic*, ed C. N. Li (New York, NY: Academic Press), 25–55.

Chen, S.-W., Wang, B., and Xu, Y. (2009). "Closely related languages, different ways of realizing focus," in *Proceedings of Interspeech 2009* (Brighton), 1007–1010.

Chen, Y., and Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *J. Phonet.* 36, 724–746. doi: 10.1016/j.wocn.2008.06.003

Christensen, R. H. B. (2019). *Ordinal - Regression Models for Ordinal Data*. R package version 2019.3-9. Available online at: https://cran.r-project.org/web/packages/ordinal/index.html

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Constant, N. (2014). *Contrastive Topic: Meanings and Realizations* (Ph.D. thesis). University of Massachusetts, Amherst, MA, United States.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Attent. Percept. Psychophys.* 20, 55–60.

Cutler, A., Dahan, A., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40, 141–201.

Cutler, A., and Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition* 7, 49–59. doi: 10.1016/0010-0277(79)90010-6

Dahan, D., Tanenhaus, M. K., and Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *J. Mem. Lang.* 47, 292–314. doi: 10.1016/S0749-596X(02)00001-3

Delin, J., and Oberlander, J. (1995). Syntactic constraints on discourse structure: the case of it-clefts. *Linguistics* 33, 465–500. doi: 10.1515/ling.1995.33.3.465

Dennison, H. Y. (2010). *Processing implied meaning through contrastive prosody* (Ph.D. thesis). University of Hawai'i at Mānoa, Hawaii, United States.

É Kiss, K. (1998). Identificational focus versus information focus. *Language* 74, 245–273. doi: 10.2307/417867

Fang, M. (1995). 汉语对比焦点的句法表现手段[The syntactic approach of Chinese contrastive focus]. 中国语文*Chinese Lang.* 4, 279–288.

Feldhausen, I., and Vanrell, M. (2015). Oraciones hendidas y marcación del foco estrecho en Español: una aproximación desde la Teoría de la Optimidad estocástica [Cleft sentences and narrow focus marking in Spanish: A stochastic Optimality Theory approach]. *Rev. Int. Lingüíst. Iberoamericana* 13, 39–59.

Féry, C., and Ishihara, S. (eds.). (2016). *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.

Féry, C., and Krifka, M. (2008). "Information structure: notional distinctions, ways of expression," in *Unity and Diversity of Languages*, ed P. V. Sterkenburg (Amsterdam: John Benjamins), 123–136.

Fodor, J. D. (1998). Learning to parse? *J. Psycholinguist. Res.* 27, 285–319.

Fodor, J. D. (2002). "Prosodic disambiguation in silent reading," in *Proceedings of North East Linguistics Society 32 (NELS32)*, ed M. Hirotani (Amherst, MA: GLSA Publications), 113–132.

Fraundorf, S. H., Benjamin, A. S., and Watson, D. G. (2013). What happened (and what did not): discourse constraints on encoding of plausible alternatives. *J. Mem. Lang.* 69, 196–227. doi: 10.1016/j.jml.2013.06.003

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *J. Mem. Lang.* 63, 367–386. doi: 10.1016/j.jml.2010.06.004

Ginzburg, J. (1994). "An update semantics for dialogue," in *Proceedings of the first International Workshop on Computational Semantics*, eds H. Bunt, R. Muskens, and G. Rentier (Tilburg: ITK, Tilburg University), 111–120.

Gotzner, N. (2017). *Alternative Sets in Language Processing*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-52761-1

Gotzner, N., Wartenburger, I., and Spalek, K. (2016). The impact of focus particles on the recognition and rejection of contrastive alternatives. *Lang. Cogn.* 8, 59–95. doi: 10.1017/langcog.2015.25

Hedberg, N. (2013). "Multiple focus and cleft sentences," in *Cleft Structures*, number 208 in Linguistik Aktuell, eds K. Hartmann, and T. Veenstra (Amsterdam; Philadelphia, PA: John Benjamins), 227–250.

Hole, D. (2011). The deconstruction of Chinese shì...de clefts revisited. *Lingua* 121, 1707–1733. doi: 10.1016/j.lingua.2011.07.004

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometr. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Husband, E. M., and Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Lang. Cogn. Neurosci.* 31, 217–235. doi: 10.1080/23273798.2015.1083113

Ip, M. H. K., and Cutler, A. (2017). "Intonation facilitates prediction of focus even in the presence of lexical tones," in *Proceedings of Interspeech 2017: Situated Interaction* (Stockholm), 1218–1222.

Ito, K., and Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J. Mem. Lang.* 58, 541–573. doi: 10.1016/j.jml.2007.06.013

Jun, S.-A. (2010). The implicit prosody hypothesis and overt prosody in English. *Lang. Cogn. Process.* 25, 1201–1233. doi: 10.1080/01690965.2010.503658

Jun, S.-A., and Bishop, J. (2015). Priming implicit prosody: prosodic boundaries and individual differences. *Lang. Speech* 58, 459–473. doi: 10.1177/0023830914563368

Kember, H., Choi, J., and Cutler, A. (2016a). "Processing advantages for focused words in Korean," in *Proceedings of Speech Prosody 2016*, eds J. Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux (Boston, MA), 702–705.

Kember, H., Choi, J., and Yu, J. (2016b). "Searching for importance: focus facilitates memory for words in English," in *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology*, eds C. Carignan and M. D. Tyler (Sydney, NSW), 181–184.

Kim, C. S. (2012). *Generating alternatives: interpreting focus in discourse* (Ph.D. thesis). University of Rochester, New York, United States.

Kim, C. S., Gunlogson, C., Tanenhaus, M. K., and Runner, J. T. (2015). Context-driven expectations about focus alternatives. *Cognition* 139, 28–49. doi: 10.1016/j.cognition.2015.02.009

Krifka, M. (2008). Basic notions of information structure. *Acta Linguist. Hungarica* 55, 243–276. doi: 10.1556/ALing.55.2008.3-4.2

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., and Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition* 133, 335–342. doi: 10.1016/j.cognition.2014.05.017

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics* 39, 463–516. doi: 10.1515/ling.2001.021

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *J. Stat. Softw.* 69, 1–33. doi: 10.18637/jss.v069.i01

Long, F. (1985). "北京话语流中声韵调的时长[Duration of initials, finals and tones in Beijing dialect]," in 北京语音实验录*[Notes of phonetic experiments in Beijing dialect]*, eds T. Lin and L. Wang (北京Beijing: 北京大学出版社Beijing University Publishing), 131–195.

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106:226.

Norris, D., Cutler, A., McQueen, J. M., and Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cogn. Psychol.* 53, 146–193. doi: 10.1016/j.cogpsych.2006.03.001

Paul, W., and Whitman, J. (2008). Shi … de focus clefts in Mandarin Chinese. *Linguist. Rev.* 25, 413–451. doi: 10.1515/TLIR.2008.012

Prince, E. F. (1978). A comparison of wh-clefts and it-clefts in discourse. *Language* 54, 883–906. doi: 10.2307/413238

Qualtrics (2017). *Qualtrics [software]*. Utah: Qualtrics.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Repp, S., and Drenhaus, H. (2015). Intonation influences processing and recall of left-dislocation sentences by indicating topic vs. focus status of dislocated referent. *Lang. Cogn. Neurosci.* 30, 324–346. doi: 10.1080/23273798.2014.923104

Riester, A. (2018). "Annotation guidelines for Questions under Discussion and information structure," in *Information Structure in Lesser-Described Languages: Studies in Syntax and Prosody*, eds E. Adamou, K. Haude, and M. Vanhove (Amsterdam: Benjamins), 403–443.

Roberts, C. (1996). "Information structure in discourse: towards an integrated formal theory of pragmatics," in *OSU Working Papers in Linguistics*, eds J. H. Yoon and A. Kathol (Columbus, OH: The Ohio State University),49. Revised version in *Semantics and Pragmatics*, 5, 1-69, 2012.

Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semant.* 1, 75–116. doi: 10.1007/BF02342617

Simpson, A. and Wu, Z. (2002). From D to T–Determiner incorporationand the creation of tense. *J. East Asian Linguist.* 11, 169–209. doi: 10.1023/A:1014934915836

Singer, M. (1976). Thematic structure and the integration of linguistic information. *J. Verbal Learn. Verbal Behav.* 15, 549–558. doi: 10.1016/0022-5371(76)90049-9

Spalek, K., Gotzner, N., and Wartenburger, I. (2014). Not only the apples: focus sensitive particles improve memory for information-structural alternatives. *J. Mem. Lang.* 70, 68–84. doi: 10.1016/j.jml.2013.09.001

Stalnaker, R. (1974). "Pragmatic presuppositions," in *Semantics and Philosophy*, eds M. K. Munitz and P. Unger (New York, NY: New York University Press), 197–213.

Stolterfoht, B., Friederici, A. D., Alter, K., and Steube, A. (2007). Processing focus structure and implicit prosody during reading: differential ERP effects. *Cognition* 104, 565–590. doi: 10.1016/j.cognition.2006.08.001

Swinney, D. A., Onifer, W., Prather, P., and Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Mem. Cogn.* 7, 159–165. doi: 10.3758/BF03197534

Vallduví, E. (2016). "Information structure," in *The Cambridge Handbook of Formal Semantics*, eds M. Aloni and P. Dekker (Cambridge: Cambridge University Press), 728–755.

Wang, B. and Xu, Y. (2006). "Prosodic encoding of topic and focus in Mandarin," in *Proceedings of Speech Prosody 2006*, eds R. Hoffmann and H. Mixdorff (Dresden: TUD Press), 313–316.

Watson, D. G., Tanenhaus, M. K., and Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H*. *Cogn. Sci.* 32, 1232–1244. doi: 10.1080/03640210802138755

Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: eye-tracking evidence for the role of contrastive accents. *Lang. Speech* 49, 367–392. doi: 10.1177/00238309060490030301

Wurm, L. H., and Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *J. Mem. Lang.* 72, 37–48. doi: 10.1016/j.jml.2013.12.003

Xie, Z. (2012). The modal uses of de and temporal shifting in Mandarin Chinese. *J. East Asian Linguist.* 21, 387–420. doi: 10.1007/s10831-012-9093-8

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *J. Phonet.* 27, 55–105.

Xu, Y. (2013). "ProsodyPro - A Tool for Large-scale Systematic Prosody analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (Aix-en-Provence), 7–10.

Yan, M. (2017). "The role of context in association strength and word retrieval," in *The 2017 Conference of the Linguistics Society of New Zealand* (Auckland).

Yan, M., Calhoun, S., and Warren, P. (2019). "The role of prosody in priming alternatives in Mandarin Chinese," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, eds S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Canberra, ACT: Australasian Speech Science and Technology Association), 437–441.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

OPINION
published: 03 September 2019
doi: 10.3389/fcomm.2019.00047

Exhaustification and Contextual Restriction

*Tue Trinh**

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany

Keywords: implicatures, alternatives, exhaustification, salience, atomicity

1. EXHAUSTIFICATION

1.1 Scientific theories are formulated based on experiments and our hunch about what must be right, i.e., "conceptual necessities" or "truisms." For the study of language, these truisms include the thesis that linguistic communication is cooperative: speakers stick to the point and tell "the truth, the whole truth, and nothing but the truth" (Fox, 2016). Here is one way to spell this out[1].

(1) Cooperative Speaker (CS)
Given a question under discussion Q, a speaker x will, by default, assert a proposition ϕ such that
(i) $\phi \in R_Q$ — "ϕ is relevant with respect to Q"
(ii) $K(\phi)$ — "x believes ϕ"
(iii) $\forall\psi\big((\psi \in R_Q \wedge K(\psi)) \rightarrow \phi \subseteq \psi\big)$ — "ϕ entails every relevant proposition that x believes"

CS turns out to be at odds with some of the most common observations. Here is one, for example. Suppose Q is who John talked to, with "who" ranging over Mary and Sue. Thus, M = [John talked to Mary] and S = [John talked to Sue] are both relevant. Assuming that relevance with respect to some question under discussion is closed under Boolean operations (cf. Groenendijk and Stokhof, 1984; Lewis, 1988),

(2) Closure conditions on relevance
(i) If ϕ is relevant, $\neg\phi$ is relevant
(ii) If ϕ and ψ are relevant, $\phi \wedge \psi$ is relevant

the set R_Q of relevant sentences with respect to Q will be $BC(\{M, S\})$, i.e., the Boolean closure of $\{M, S\}$[2]. Now let us say that the strongest proposition in this set entailed by the speaker's belief is $M \wedge \neg S$. From CS it follows that the speaker will assert not M, not $\neg S$, but $M \wedge \neg S$, the strongest of the three, which means she will utter not (3a), not (3b), but (3c).

(3) a. John talked to Mary
b. John didn't talk to Sue
c. John talked to Mary but not Sue

The fact, however, is that the speaker can freely choose between (3a) and (3c). The discrepancy between prediction and observation is shown in (4).

[1] Clauses (i), (ii) and (iii) of CS derive from the Gricean maxims of Relation, Quality, and Quantity, respectively (Grice, 1967). There is a fourth maxim, Manner, which is orthogonal to what we will be discussing.

[2] This means being relevant with respect to Q is not the same as being a congruent answer of Q. For example, "John did not talk to Mary" is relevant to but not a congruent answer of "who did John talk to." I thank one of the reviewers for drawing my attention to this distinction.

OPEN ACCESS

Edited by:
Sophie Repp,
University of Cologne, Germany

Reviewed by:
Alexandre Cremers,
University of Amsterdam, Netherlands
Edgar Onea Onea,
University of Graz, Austria

***Correspondence:**
Tue Trinh
tuetrinh@alum.mit.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *11 March 2019*
Accepted: *15 August 2019*
Published: *03 September 2019*

Citation:
Trinh T (2019) Exhaustification and Contextual Restriction. Front. Commun. 4:47. doi: 10.3389/fcomm.2019.00047

(4) Speaker's belief: $M \wedge \neg S$

	predicted	observed
John talked to Mary	✗	✓
John didn't talk to Sue	✗	✗
John talked to Mary but not Sue	✓	✓

Thus, two sentences which are predicted to have different meanings behave as if they have the same meaning under the assumption that CS is true.

Here is another example. Suppose the strongest relevant proposition entailed by the speaker's belief is M. Given CS, she will assert M. Since M and $M \vee (M \wedge S)$ are equivalent, we predict that the speaker can freely choose between (5a) and (5b).

(5) a. John talked to Mary
b. John talked to Mary, or both Mary and Sue

The fact, however, is that she will say (5b), not (5a). The discrepancy between prediction and observation is shown in (6).

(6) Speaker's belief: M

	predicted	observed
John talked to Mary	✓	✗
John talked to Mary, or to both Mary and Sue	✓	✓

Thus, two sentences which are predicted to have the same meaning behave as if they have different meanings under the assumption that CS is true.

1.2 Paul Grice sketches a way to derive such facts from CS itself plus other assumptions about how people think when they speak (Grice, 1967). The goal is to not tinker with the grammar[3]. Here is his account of the first example. Suppose the speaker asserts M and nothing else. It follows, from CS, that K(M), and given that S is relevant and not entailed by M, it follows, again from CS, that $\neg$K(S). Assuming that the speaker is opinionated about S, i.e., that $K(S) \vee K(\neg S)$, it follows that $K(\neg S)$. And from K(M) and $K(\neg S)$ it follows that $K(M \wedge \neg S)$[4]. Thus, by asserting M, the speaker conveys $K(M \wedge \neg S)$, which would also be conveyed by her asserting $M \wedge \neg S$. This is why she can freely choose between (3a) and (3c): these sentences convey the same belief, namely $M \wedge \neg S$. To introduce some terminology at this point, M is the "assertion," S the "alternative," $\neg S$ the "scalar implicature," and $M \wedge \neg S$ the "strengthened meaning."

This account is appealing. It accords with our intuition that messages can be conveyed by what is said plus reasoning about what could have been said. All we must do is change "assert" to "convey" in the definition of CS, which we are, of course, willing to do. As it turns out, however, Grice's account is flawed. The crack is R_Q. Recall that this set is assumed to be closed under Boolean operations. Thus, it contains both S and $\neg S$, which means the same reasoning as presented above could be applied with $\neg S$ in place of S, and the speaker would convey, by asserting M, not $K(M \wedge \neg S)$, but $K(M \wedge S)$. The reasoning would go like this. Suppose the speaker asserts M and nothing else. It follows, from CS, that K(M), and given that $\neg S$ is relevant and is not entailed by M, it follows, again from CS, that $\neg K(\neg S)$. Assuming that the speaker is opinionated about $\neg S$, i.e., that $K(\neg S) \vee K(\neg\neg S)$, it follows that $K(\neg\neg S)$, i.e., that K(S). And from K(M) and K(S) it follows that $K(M \wedge S)$.

Grice's attempt at explaining the first example failed. The reason is that there are too many alternatives in R_Q. Specifically, there are those which give rise to incompatible strengthened meanings. These are called "symmetric alternatives," and the problem they pose is called the "symmetry problem[5]". An obvious fix, therefore, is to tamper with R_Q, specifically to "break" the symmetry in this set by keeping S in it and pruning $\neg S$ from it. But this amounts, practically, to being incoherent: how can a proposition be relevant while its negation is not[6]?

In addition, Grice has no explanation for the second example: even if $\neg S$ could be pruned from R_Q, both (5a) and (5b) should still express the proposition M, which means asserting either of them should license the same implicature $\neg S$, clearly not what is observed[7].

1.3 The "grammatical approach to implicatures" provides a way to square CS with the facts (cf. Krifka, 1995; Fox, 2007; Chierchia et al., 2012). The idea is that we have to tinker with the grammar after all. The core of the proposal is the postulation of a covert lexical item, EXH_C, which composes with a sentence ϕ, its "prejacent," to affirm ϕ and negate a selection of ϕ's alternatives[8].

(7) Interpretation of EXH_C

$$[EXH_C\ \phi] \Leftrightarrow \phi \wedge \bigwedge \left\{\neg\psi \mid \psi \in E_\phi\big(F(\phi) \cap C\big)\right\}$$

Procedurally, the calculation of $[EXH_C\ \ \phi]$ consists of the following steps: (i) pick a set C of "contextually salient" alternatives; (ii) restrict C by intersecting it with the set $F(\phi)$ of "formal" alternatives; (iii) select from $F(\phi) \cap C$ the set $E_\phi\big(F(\phi) \cap C\big)$ of "innocently excludable" alternatives; (iv) conjoin ϕ with the negation of every innocently excludable alternative.

The value of C is discourse dependent. A natural candidate for C is R_Q, the set of relevant sentences given the question under discussion Q. The functions F and E_ϕ are defined in syntactic and semantic terms, respectively[9].

[3]As correctly pointed out by one of the reviewers, the "Gricean account" presented here is not literally Grice (1967)—that work did not formulate CS or impose closure conditions on relevance, for example—but an exegesis thereof which has become more or less established among those who work in this area. The reformulation of the maxims of conversation as CS in (1) (see note 1) and the explication of relevance as subject to the closure conditions in (2) can be considered amendments which are compatible with what Grice says and at the same time sharpen it into a proposal that can be more concretely evaluated and critiqued.

[4]The assumption of an "opinionated speaker" was not made explicit in Grice (1967). That it must be added to strengthen $\neg$K(S) to the more observationally adequate $K(\neg S)$ was noted by others (cf. Soames, 1982; Horn, 1989; Sauerland, 2004). See note 3.

[5]Instances of the symmetry problem were first pointed out in Kroch (1972).

[6]We can think of a relevant proposition as one whose truth value we are interested in finding out. Obviously, we cannot want to know whether p is true without wanting to know whether $\neg p$ is true.

[7]The problem of semantically equivalent sentences licensing different implicatures is sometimes called the "functionality problem" (Van Rooij and Schulz, 2004).

[8]The definition in (7) was proposed in Fox (2007). I write "$\bigwedge X$" for the conjunction of all elements of X.

[9]Thus, EXH_C shares with pronominal elements in having an indexical component in addition to syntactic and semantic features. Take the pronoun $[him_7]$, for

(8) Definition of formal and innocently excludable alternatives[10].

a. $F(\phi) = \{\psi \mid \psi$ is no more complex than $\phi\}$
b. $E_\phi(X) := \bigcap\{X' \mid X'$ is a maximal subset of X such that $\{\phi\} \cup \{\neg\psi \mid \psi \in X'\}$ is consistent$\}$

Katzir (2007), and later Fox and Katzir (2011), propose to explicate the relation "is no more complex than" in (8a) as follows: ψ is no more complex than ϕ iff ψ is derivable from ϕ by successively replacing constituents of sentences, beginning with ϕ itself, with elements of the "substitution source" (SS), defined as in (9).

(9) Substitution source (SS)
SS = {x | x is a lexical item} ∪ {x | x is a constituent of an expression uttered in the context}

As for (8b), the idea is this: (i) try to build maximal subsets of X which contain sentences that can be consistently negated together with ϕ, then intersect these subsets: in this intersection are the innocently excludable alternatives.

To illustrate how the system works, suppose a disjunction of M and S was uttered, with (10) as its parse.

(10) $\text{EXH}_C\ [_\phi\ M \vee S]$

Given that ϕ, M and S have all been uttered, from the prejacent ϕ we can derive ϕ by replacing ϕ with itself, M by replacing ϕ with M, S by replacing ϕ with S, and $[M \wedge S]$ by replacing $\vee$ with $\wedge$. Assuming $C = BC(\{M, S\})$, we have $F(\phi) \cap C = \{\phi, M, S, M \wedge S\}$. Call this set A_1. Let us now try to build maximal subsets of A_1 which contain sentences that can be consistently negated together with ϕ. One such set is $A_2 = \{M, M \wedge S\}$, since $\{\phi, \neg M, \neg[M \wedge S]\}$ is consistent. Another is $A_3 = \{S, M \wedge S\}$, since $\{\phi, \neg S, \neg[M \wedge S]\}$ is consistent. Now intersect A_2 and A_3. The result is $A_4 = A_2 \cap A_3 = \{M \wedge S\}$. This is the set of innocently excludable alternatives. We derive (11).

(11) $\text{EXH}_C\ [_\phi\ M \vee S] \Leftrightarrow \phi \wedge \neg[M \wedge S]$

This explains the availability of the "exclusive" reading of plain disjunctions. For example, the string in (12a) will now have (12b) as parse, which is interpreted as $(M \vee S) \wedge \neg(M \wedge S)$, i.e., as "John talked to Mary or Sue but not both."

(12) a. John talked to Mary or Sue
b. $\text{EXH}_{BC(\{M, S\})}$ [John talked to Mary or Sue]

In the above example, the individual disjuncts derived by F from the disjunction end up being excluded from the domain of exhaustification by E_ϕ. The situation changes when the disjunction is embedded under a universal quantifier, such as the modal necessity operator, as in (13) (Sauerland, 2004)[11].

(13) $\text{EXH}_C\ [_\phi\ \Box[_\psi\ M \vee S]]$

Applying F to the prejacent ϕ, we derive, again, ϕ by replacing ϕ with itself, $\Box M$ by replacing ψ with M, $\Box S$ by replacing ψ with S, and $\Box[M \wedge S]$ by replacing $\vee$ with $\wedge$. Assuming $C = BC(\{\Box M, \Box S\})$, this means $F(\phi) \cap C = \{\phi, \Box M, \Box S, \Box[M \wedge S]\}$, which means $E_\phi(F(\phi) \cap C) = \{\Box M, \Box S, \Box[M \wedge S]\}$. We derive (14).

(14) $\text{EXH}_C\ [_\phi\ \Box[_\psi\ M \vee S]] \Leftrightarrow \phi \wedge \neg\Box M \wedge \neg\Box S \Leftrightarrow \phi \wedge \Diamond M \wedge \Diamond S$

This explains the availability of the "distributive" reading of disjunctions embedded under universal quantifiers. For example, the string in (15a) will now have (15b) as parse, which is interpreted as "John is required to talk to Mary or Sue and he is allowed to talk to Mary and he is allowed to talk to Sue," exactly what is observed.

(15) a. John is required to talk to Mary or Sue
b. $\text{EXH}_{BC(\{\Box M, \Box S\})}$ □[John talked to Mary or Sue]

Note that F can also derive alternatives by replacing constituents of the prejacent with linguistic materials which have been used in the discourse context but which are not part of the prejacent itself. This makes it possible to analyze "particularized implicatures" as cases of exhaustification. As an example, consider the inference licensed by A's response to B's question below (cf. Matsumoto, 1995; Katzir, 2007).

(16) A: Yesterday it was warm and sunny with gusts of wind.
B: What about today?
A: Today it was warm.

A's response to B's question licenses the implicature that today it was not warm and sunny with gusts of wind. The alternative needed for this implicature is derived from the prejacent by replacing [warm] with [warm and sunny with gusts of wind], which is a constituent that has been uttered in the context but is not part of the prejacent itself[12].

Coming back to the two apparent counter examples to CS discussed 1.1, here is what the grammatical approach to implicatures can say. Let us start with the first one. The puzzle posed by this example can be formulated in terms of this question: how can (17a) and (17b) be equivalent?

(17) a. John talked to Mary
b. John talked to Mary but not Sue

example. It carries an index which is assigned a value by the context. Semantically, it refers to male entities. Syntactically, it cannot be bound within the smallest clause, etc.

[10]Notationally, $\bigcap X$ is the intersection of all elements of X.

[11]Following standard practice, I use □ to represent universal modals, such as [must], [have to], [be required], etc., and ◇ to represent existential modals, such as [may], [can], [be allowed], etc.

[12]In text object language expressions are put inside square brackets.

The introduction of EXH_C into the lexicon makes it possible to understand this question not as rhetorical, but technical. It now has a straightforward answer: (17a) can be parsed as (18).

(18) EXH_C [$_\phi$ John talked to Mary]
where C = BC({John talked to Mary, John talked to Sue})

As it turns out, among the members of C, only (19a), besides ϕ itself, is no more complex than the prejacent: (19a) is derivable from ϕ by replacing [Mary] with [Sue], but no other member of C, besides ϕ itself, is derivable from ϕ by successively replacing constituents of ϕ with salient linguistic materials or items stored in the lexicon. Crucially, (19b) is not derivable from ϕ in that way.

(19) a. John talked to Sue
b. John did not talk to Sue

Given $C = BC(\{M, S\})$, we have $F(\phi) \cap C = \{\phi, (19a)\}$, which means $E_\phi(F(\phi) \cap C) = \{(19a)\}$, which means [$EXH_C$ [$_\phi$ John talked to Mary]] $\Leftrightarrow \phi \wedge \neg$[John talked to Sue]. This is how (17a) and (17b) can be equivalent. Note, however, that explaining how (17a) and (17b) can be equivalent actually falls short of accounting for the facts. In reality, these sentences not only can, but must be equivalent: if (17a) is uttered where $R_Q = BC(\{M, S\})$, it will have to be understood as $M \wedge \neg S$. This means, in effect, that a parse with EXH_C is the default (cf. Krifka, 1995; Fox, 2007; Magri, 2009). The generalization is stated in (20).

(20) Mandatory Exhaustification
Every matrix sentence is parsed with EXH_C by default

Let us now consider the second example. The problem posed by this example can be formulated in terms of the following question: how can (21a) and (21b) not be equivalent?

(21) a. John talked to Mary
b. John talked to Mary, or both Mary and Sue

Just as for the first example, the answer is now straightforward. Keeping to $C = BC(\{M, S\})$, the following two parses will deliver the right result. Note that the outer EXH_C in (22b) is merged to satisfy (20). Semantically, it is vacuous.

(22) a. EXH_C [John talked to Mary]
b. EXH_C [[EXH_C John talked to Mary] or [John talked to Mary and Sue]]

We already know that (22a) means $M \wedge \neg S$. The reader is invited to verify for herself that (22b) means $M \vee (M \wedge S)$, which is equivalent to M. Note that there are two instances of EXH_C in (22b), one of which is embedded. If EXH_C is merged at only the matrix level, as in (23), the resulting meaning will be $M \wedge \neg S$, as can also be verified by the reader.

(23) EXH_C [[John talked to Mary] or [John talked to Mary and Sue]]

The embeddability of EXH_C corroborates the claim that it is a lexical item which can be syntactically integrated, not a notational device which models pragmatic reasoning performed at the speech act level (cf. Fox, 2007; Magri, 2009; Chierchia et al., 2012; Crnič, 2012).

But we again fall short of explaining the phenomenon: it is not that (21b) can be interpreted as M, but that it must be. Specifically, (21b) cannot be interpreted as $M \wedge \neg S$. This means that (21b) must be parsed as (22b) and cannot be parsed as (23). Obviously, (20) does not explain this fact, as this principle only requires EXH_C at the matrix level. Note, also, that (23) is stronger than (22b). What can force a sentence to be weakened by an instance of embedded EXH_C? The answer has to do with the fact that (21b) is not deviant in the same way (24) is.

(24) #John saw a dog or an animal

This sentence violates the following principle, which for present purposes we take to be a primitive of grammar (Hurford, 1974).

(25) Hurford's Constraint (HC)
A disjunction is deviant if one disjunct entails the other

It is HC which forces (21b) to be parsed as (22b): because the first disjunct is exhaustified, it is incompatible with the second disjunct. The parse in (23), on the other hand, violates HC, since the second disjunct of the prejacent entails the first[13].

1.4. At this point, an irony is perhaps worth noting: CS was originally motivated by the wish to derive implicatures from extra-grammatical principles, but it turns out that in order to keep CS, implicatures have to be derived in the grammar. The question to ask is what role CS plays in the interpretation of utterances. The answer is that CS derives "ignorance inferences." A simple consequence of CS is (26).

(26) Ignorance inference
A speaker who asserts ϕ is ignorant about every relevant ψ which is not settled by ϕ

We say that the speaker is "ignorant" about ϕ iff $\neg K(\phi) \wedge \neg K(\neg\phi)$, and say that ϕ "settles" ψ iff ϕ entails ψ or ϕ entails $\neg\psi$. Here, then, is how (26) follows from CS. Given CS, the speaker's assertion should entail every relevant proposition which she believes to be true. Suppose ψ is relevant. Then, $\neg\psi$ is relevant also. If the speaker's assertion entails neither ψ nor $\neg\psi$, then it must be that she believes neither ψ nor $\neg\psi$, i.e., that she is ignorant about ψ.

The theorem in (26) is easy to illustrate. Consider, again (21b), which is repeated in (27a). As we have argued, this sentence is parsed as (27b).

(27) a. John talked to Mary, or both Mary and Sue
b. EXH_C [[EXH_C John talked to Mary] or [John talked to Mary and Sue]]

Observationally, (27a) conveys the message that the speaker is ignorant about the proposition that John talked to Sue: if asked whether John talked to Sue, she will not be able to say yes or say no truthfully. In other words, (27a) conveys the message that $\neg K(S) \wedge \neg K(\neg S)$. This is predicted, since it follows from (27b) neither that John talked to Sue, nor that John did not talk to Sue.

[13] See Meyer (2013, 2014) for an explanation as to why exhaustification cannot rescue (24).

2. CONTEXTUAL RESTRICTION

2.1 Recall Grice's problem: there are too many alternatives in R_Q. Specifically, for every alternative in R_Q which could give rise to the attested strengthened meaning, there is a "symmetric" counterpart in the same set which could give rise to a non-attested strengthened meaning inconsistent with the attested one. Let us state the sense of "symmetric alternatives" which has underlied our usage of this notion so far.

(28) Symmetric alternatives
ψ and ψ' are symmetric alternatives of ϕ in X iff
(i) $\psi, \psi' \in X$
(ii) $\{\phi, \neg\psi\}$ and $\{\phi, \neg\psi'\}$ are consistent
(iii) $\{\phi, \neg\psi, \neg\psi'\}$ is inconsistent

Thus, S and ¬S are symmetric alternatives of M in $BC(\{M, S\})$. From the assertion of M, pragmatic reasoning based on CS will derive $K(S \wedge \neg S)$ under the assumption that the speaker is opinionated about S, and derive $\neg K(S) \wedge \neg K(\neg S)$ under the assumption that the speaker is not opinionated about S. The first inference means the speaker is incoherent, the second means she is ignorant about S, i.e., that she has no idea whether S is true or not. Neither accords with intuition, as an assertion of M is observed to license the inference that $K(\neg S)$, i.e., that the speaker believes that S is false. This discrepancy between prediction and observation which results from the existence of symmetric alternatives is the "symmetry problem." As we saw in the last section, the grammatical approach to implicatures solves this problem, or more precisely this instance of it, by denying the premise that M is the assertion. Instead, it claims that what was asserted is really $[EXH_C\ M]$, which is, by virtue of compositional semantics, synonymous with $[M \wedge \neg S]$. In this case, symmetry is broken by F: one of the symmetric alternatives, ¬S, is more complex than the prejacent, while the other, S, is not.

2.2 Let us now turn to a discussion of salience. We have identified the set of salient sentences with the set of sentences relevant to the question under discussion. We can ask whether salience is just relevance. Since the only condition on relevance, by assumption, is closure under Boolean operations, the question amounts to whether salience is the same as closure under Boolean operations, and if not, how these notions are related. One way to frame the issue is to ask what the relationship is between the sets REL and SAL, defined in (29).

(29) a. REL:= {X | X is closed under Boolean operations}
b. SAL:={X | X is contextually salient}

To be contextually salient, or in short, salient, is to be the value of C such that $[EXH_C\ \phi]$ licenses the attested inference in the given context. I will argue that there is no systematic relationship between REL and SAL. Specifically, I will argue that these sets partially overlap, i.e., that $REL - SAL$, $REL \cap SAL$ and $SAL - REL$ are all non-empty.

Let us show that $REL \cap SAL \neq \{\ \}$. Consider the discourse below, where A's response to B's utterance is parsed as indicated.

(30) A: What do you want to know?
B: I want to know who John talked to.
A: EXH_C [John talked to Mary].
$\rightsquigarrow M \wedge \neg S$

The last sentence, with the indicated parse, licenses the inference that John talked to Mary only. This inference can be derived by identifying C with the set $BC(\{M, S\})$. This set, being closed under Boolean operations, is also a member of REL, which means it is both in SAL and in REL, which means $REL \cap SAL \neq \{\ \}$.

Let us show that $REL - SAL \neq \{\ \}$. Consider the discourse below[14].

(31) A: What do you want to know?
B: I want to know whether John talked to Mary, and in case he didn't, whether he talked to Sue.
A: EXH_C [John talked to Mary or Sue].
$\not\rightsquigarrow \neg M \wedge S$

The question under discussion—"whether John talked to Mary, and in case he didn't, whether he talked to Sue"—partitions logical space into the cells $M, \neg M \wedge S, \neg M \wedge \neg S$. This partition corresponds to the set $BC(\{M \vee S, M\})$, which is an element of REL. However, it is not salient: if it were, the last sentence in (31), parsed as indicated, would license the inference that John talked to only Sue. But we observe that sentence does not license this inference. Therefore, $REL - SAL \neq \{\ \}$.

To see that $SAL - REL \neq \{\ \}$, consider the discourse in (32).

(32) A: What do you want to know?
B: I want to know how the students did on the exam.
A: EXH_C [Not all of them passed the exam].
$\rightsquigarrow$ some of them did

The last sentence in (32), with the indicated parse, licenses the inference that some of the students passed the exam. This must result from $E_\phi(F(\phi) \cap C)$ containing (33a) but not (33b).

(33) a. [not [some of the students passed the exam]]
b. [some of the students passed the exam]

As $F(\phi)$, by definition, contains both (33a) and (33b) and E_ϕ, also by definition, cannot prune one to the exclusion of the other, C must contain (33a) but not (33b), i.e., C must be the set {(33a)}. This set is not closed under Boolean operations, so it is not in REL. Given that it is in SAL, because it is the value of C, it is in $SAL - REL$, which means $SAL - REL \neq \{\ \}$.

2.3 We have just seen that there is no systematic relationship between relevance and salience: there are relevance sets which are salience sets ($REL \cap SAL \neq \{\ \}$), relevance sets which are not salience sets ($REL - SAL \neq \{\ \}$), and salience sets which are not relevance sets ($SAL - REL \neq \{\ \}$). We will now consider another possible criterion for salience: utterance. Let us define

[14] I assume that in general, any set of propositions which is closed under Boolean operations corresponds to a possible question under discussion. I thank one of the reviewers for suggesting that this assumption needs to be stated, and also, for suggesting the discourse context in (31) which makes the set $BC(\{M, M \vee S\})$ relevant.

the set UTT as containing sets of sentences which are derived by using linguistic materials that have been uttered in the context[15].

(34) UTT:={X | X ⊆ {Y | Y is derived by using linguistic materials that have been uttered in the context}}

What is the relationship between UTT and SAL? Again, I will argue that there is no systematic relationship between UTT and SAL, specifically that UTT ∩ SAL, UTT − SAL, and SAL − UTT are all non-empty.

To see both that UTT∩SAL ≠ { } and that UTT−SAL ≠ { }, consider the discourse in (35).

(35) A: John went for a run. He didn't smoke.
B: What about Bill?
A: Bill went for a run.
⇝ Bill smoked

A's response to B's question, parsed as (36), clearly has a reading which implies that Bill is not like John, i.e., that he smoked. In this case, F(ϕ) contains both (36a), derived from ϕ by replacing [$_{VP}$ go for a run] with [$_{VP}$ smoke], and (36b), derived from ϕ by replacing [$_{\overline{T}}$ went for a run] with [$_{\overline{T}}$ didn't smoke].

(36) EXH$_C$ [$_\phi$ Bill went for a run]
a. [$_\psi$ Bill smoked]
b. [$_{\psi'}$ Bill didn't smoke]

As both [$_{\overline{T}}$ didn't smoke] and [$_{VP}$ smoke] are constituents that have been uttered, UTT contains {(36a)}, {(36b)}, and {(36a), (36b)}. However, the attested inference of (36), namely that Bill smoked, requires that C contain (36b) but not (36a), i.e., that C = {(36b)}. This means both that SAL contains something which is in UTT and that UTT contains something which is not in SAL, i.e., both that UTT ∩ SAL ≠ { } and that UTT − SAL ≠ { }.

It remains to show that SAL − UTT ≠ { }. This we have actually done with the discourse in (30), where the last sentence, repeated here in (37a), is observed to license the inference that (37b) is false. This inference requires (37b) to be a formal alternative of (37a), i.e., to be derivable from (37a) by replacing [Mary] with [Sue].

(37) a. John talked to Mary
b. John talked to Sue

The context we constructed is such that [Sue] has not been uttered. The attested inference, thus, shows that (37b) is in SAL but not in UTT, i.e., that SAL − UTT ≠ { }.

2.4 Relevance and utterance can be defined with sufficient precision to make concrete predictions. However, C cannot be defined in terms of these notions, and it is, at this point, not clear what other notion can be resorted to in establishing an understanding of salience. This poses a threat to the predictive power of the grammatical approach to implicatures. Katzir (2014) suggests a way to circumvent this threat: to change the definition of EXH$_C$ from (7), repeated in (38a) to (38b). The new definition amounts to stipulating that salience, whatever it is, cannot break symmetry: it allows E$_\phi$ to weed out any symmetry in F(ϕ) before C has a chance.

(38) a. $[\text{EXH}_C\ \phi] \Leftrightarrow \phi \wedge \bigwedge \left\{\neg\psi \mid \psi \in E_\phi\big(F(\phi) \cap C\big)\right\}$
old definition

b. $[\text{EXH}_C\ \phi] \Leftrightarrow \phi \wedge \bigwedge \left\{\neg\psi \mid \psi \in C \cap E_\phi\big(F(\phi)\big)\right\}$
new definition

Katzir's proposal solves the problem posed by (31). What we want to rule out is the possibility of M ∨ S having M but not S as alternative. This is achieved by (38b): although both M and S are in F(M ∨ S), they will both be eliminated from $E_\phi\big(F(\phi)\big)$ by the definition of E$_\phi$. Thus, the unattested implicature can never arise, no matter what C is.

Katzir's solution, however, fails to account for the other cases of unpredicted symmetry breaking, namely (32) and (35). Trinh and Haida (2015) propose to deal with (32) by revising the definition of F. They suggest to impose the following condition on F, specifically on the replacement operation which derives alternatives from the prejacent.

(39) Atomicity (first part)
[u/v]([x/y](z)) is undefined if u is a subconstituent of y

where [x/y](z) stands for the result of replacing x in z with y, i.e., the result of applying the replacement of x with y to z. Call z the "input," x the "target," and y the "substitute." What atomicity says is that no target may be a subconstituent of a substitute. The condition prevents (33b) to be derived from (32). This derivation would involve two steps: (i) replacing ϕ with ψ and (ii) replacing [all], a subconstituent of ψ, with [some]. The second step is ruled out by Atomicity: [all/some]([ϕ/ψ](ϕ)) is undefined, because [all] is a subconstituent of ψ[16].

(40) [$_\phi$ not [$_\psi$ all of the students passed the exam]]

Trinh (2018) proposes to add another clause to Atomicity to deal with the case in (35). This clause constrains what can be a substitution source.

(41) Atomicity (second part)
If x, y ∈ SS and neither x nor y is a lexical item, x is not a subconstituent of y

This condition would prevent [$_{VP}$ smoke] and [$_{\overline{T}}$ didn't smoke] from both being elements of SS, thus preventing [$_\psi$ Bill smoked] and [$_{\psi'}$ Bill didn't smoke] from both being alternatives of [$_\phi$ Bill went for a run]. It would, however, allow [$_{\psi'}$ Bill didn't smoke] as an alternative, making it possible to derive the attested implicature[17].

[15]Naturally, "having been uttered" cannot be taken to mean "consisting of morphemes that have been pronounced." So much of language is silent that a collection of overt morphemes will rarely yield any interpretation. To "utter X" has to be understood as to "give some hints about X by way of making sound." In that sense, elliptical sentences, copies of movement, null pronouns, etc., are all part of the utterance.

[16]Note that (33b) cannot be derived from (32) by first replacing [all] in (32) with [some], generating [$_\alpha$ not [$_\beta$ some of the students passed the exam]], followed by replacing α with β. The second step of this derivation is by definition not legitimate, since β is not an element of SS: it is not a lexical item, and it is not a constituent of any expression uttered in the context.

Apart from being a partial solution which necessitates the complication of F, the new definition of EXH_C also raises a "why" question: why does language opt for (38b) instead of (38a)? There is a certain conceptual naturalness to the presence of C, F, and E_ϕ in the definition of exhaustification, since these mean pragmatic, syntactic, and semantic factors are all involved, a common feature of linguistic interpretation. However, it is much harder, if possible at all, to say why the order of operations in (38b) is more natural than that in (38a)[18].

3. CONCLUSION

The grammatical approach to implicatures derives what was traditionally considered "pragmatic" inferences in the grammar by way of compositional semantics. It succeeds in accounting for inference patterns exhibited by a large number of cases while keeping to an overwhelmingly intuitive understanding of language use. But the proposal has a black box: the predictions it makes depend on which sentences are salient. It turns out, on closer inspection, that there is no systematic relationship between salience and relevance, nor is there one between salience and utterance. One way to resolve this problem is to stipulate that salience cannot break symmetry, and on that stipulation, calibrate other components of exhaustification to account for the facts. This strategy necessitates the complication of F, the function which maps sentences to their structurally simpler alternatives. Also, it brings into clear relief the need for a deeper explanation for the way pragmatics, semantics, and syntax interact in the process of exhaustification.

[17]See Trinh (2018) for arguments as to why [smoke] should be considered a VP and not a lexical item. Also, note that (41) will not prevent the derivation of the bipartite conjunction [B and C] from the tripartite disjunction [A or [B or C]], and thus, will not prevent us from deriving the "only one" interpretation of tripartite disjunctions. This is because [B and C] is equivalent to [B and [C or C]], etc. In other words, the only non-lexical members which SS must contain are the individual disjuncts A, B, and C, none of which is a subconstituent of any other. I thank one of the reviewers for raising this point.

[18]One is reminded of the issue raised by Mats Rooth about Irene Heim's definition of the context change potentials of logical connectives: there are several definitions of, say, [if A, B] which gets the truth conditions right, but only one which gets the projection of presuppositions right, and there is no clear reason why one definition is more natural than the others (cf. Heim, 1990; Schlenker, 2008).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was funded by the ERC Advanced Grant Speech Acts in Grammar and Discourse (SPAGAD), ERC-2017-ADG 787929. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

REFERENCES

Chierchia, G., Fox, D., and Spector, B. (2012). "The grammatical view of scalar implicatures and the relationship between semantics and pragmatics," in *Semantics: An International Handbook of Natural Language Meaning*, eds P. Portner, C. Maienborn, and K. von Heusinger (Berlin: De Gruyter), 2297–2232.

Crnič, L. (2012). Focus particles and embedded exhaustification. *J. Semant.* 30, 533–558. doi: 10.1093/jos/ffs018

Fox, D. (2007). "Free choice disjunction and the theory of scalar implicatures," in *Presupposition and Implicature in Compositional Semantics*, eds U. Sauerland, and P. Stateva (New York, NY: Palgrave-Macmillan), 71–120.

Fox, D. (2016). *Topics in Exhaustification*. Minicourse taught at the Hebrew University of Jerusalem, Jerusalem.

Fox, D., and Katzir, R. (2011). On the characterization of alternatives. *Nat. Lang. Semant.* 19, 87–107. doi: 10.1007/s11050-010-9065-3

Grice, P. (1967). *Logic and Conversation*. William James Lectures. Cambridge: Harvard University Press.

Groenendijk, J., and Stokhof, M. (1984). *Studies on the semantics of questions and the pragmatics of answers* (PhD thesis), University of Amsterdam, Amsterdam, Netherlands.

Heim, I. (1990). "Presupposition projection," in *Presupposition, Lexical Meaning and Discourse Processes: Workshop Reader*, ed R. van der Sandt (Nijmegen: University of Nijmegen).

Horn, L. (1989). *A Natural History of Negation*. Chicago, IL: University of Chicago Press.

Hurford, J. R. (1974). Exclusive or inclusive disjunction. *Found. Lang.* 11, 409–411.

Katzir, R. (2007). Structurally-defined alternatives. *Linguist. Philos.* 30, 669–690. doi: 10.1007/s10988-008-9029-y

Katzir, R. (2014). "On the roles of markedness and contradiction in the use of alternatives," in *Semantics, Pragmatics and the Case of Scalar Implicatures*, ed S. Pistoia-Reda (London: Palgrave), 40–71.

Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguist. Anal.* 25, 209–257. doi: 10.3765/salt.v4i0.2462

Kroch, A. (1972). Lexical and inferred meanings for some time adverbials. *Q. Progr. Rep. Res. Lab. Electron.* 104, 260–267.

Lewis, D. (1988). Relevant implication. *Theoria* 54, 161–174. doi: 10.1111/j.1755-2567.1988.tb00716.x

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Nat. Lang. Semant.* 17, 245–297. doi: 10.1007/s11050-009-9042-x

Matsumoto, Y. (1995). The conversational condition on horn scales. *Linguist. Philos.* 18, 21–60. doi: 10.1007/BF00984960

Meyer, M.-C. (2013). *Ignorance and grammar* (PhD thesis), Massachusetts Institute of Technology, Cambridge, MA, United States.

Meyer, M. C. (2014). "Deriving Hurford's constraint," in *Proceedings of SALT 24* (New York, NY), 577–596.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguist. Philos.* 27, 367–391. doi: 10.1023/B:LING.0000023378.71748.db

Schlenker, P. (2008). Be articulate: a pragmatic theory of presupposition projection. *Theor. Linguist.* 34, 157–212. doi: 10.1515/THLI.2008.013

Soames, S. (1982). How presuppositions are inherited: a solution to the projection problem. *Linguist. Inq.* 13, 483–545.

Trinh, T. (2018). Keeping it simple. *Nat. Lang. Semant.* 26, 111–124. doi: 10.1007/s11050-018-9143-5

Trinh, T., and Haida, A. (2015). Constraining the derivation of alternatives. *Nat. Lang. Semant.* 23, 249–270. doi: 10.1007/s11050-015-9115-y

Van Rooij, R., and Schulz, K. (2004). Exhaustive interpretation of complex sentences. *J. Logic Lang. Inform.* 13, 491–519. doi: 10.1007/s10849-004-2118-6

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 September 2019
doi: 10.3389/fpsyg.2019.01993

Prenuclear L*+H Activates Alternatives for the Accented Word

Bettina Braun[1]* *and María Biezma*[2]

[1] *Department of Linguistics, University of Konstanz, Konstanz, Germany,* [2] *Spanish and Portuguese Department, University of Massachusetts Amherst, Amherst, MA, United States*

Previous processing studies have shown that constituents that are prosodically marked as focus lead to an activation of alternatives. We investigate the processing of constituents that are prosodically marked as *contrastive topics*. In German, contrastive topics are prosodically realized by prenuclear L*+H accents. Our study tests (a) whether prenuclear accents (as opposed to nuclear accents) are able to activate contrastive alternatives, (b) whether they do this in the same way as constituents prosodically marked as focus with nuclear accents do, which is important for semantic modeling, and (c) whether the activation of alternatives is caused by pitch accent type (prenuclear L*+H as contrastive accent vs. prenuclear L+H* as non-contrastive accent) or by differences in F0-excursion (related to prominence). We conducted two visual-world eye-tracking studies, in which German listeners heard declarative utterances (e.g., *The swimmer wanted to put on flappers*) and watched displays that depicted four printed words: one that was a contrastive alternative to the subject noun (e.g., *diver*), one that was non-contrastively related to it (e.g., *sports*), the object (e.g., *flappers*), which had to be clicked, and an unrelated distractor. Experiment 1 presented participants with two naturally produced intonation conditions, a broad focus control condition with a prenuclear L+H* accent on the subject and a contrastive topic condition with a prenuclear L*+H accent. The results showed that participants fixated more on the contrastive alternative when the subject was produced with an L*+H accent, with the same effect size and timing as reported for focus constituents. Experiment 2 resynthesized the stimuli so that peak height and F0-excursion were the same across intonation conditions. The effect was the same, but the time course was slightly later. Our results suggest that prenuclear L*+H immediately leads to the activation of alternatives during online processing, and that the F0-excursion of the accent lends little. The results are discussed with regard to the processing of contrastive focus accents and theories of contrastive topic.

Keywords: intonation, processing, contrastive topics, alternative sets, German

OPEN ACCESS

Edited by:
Sophie Repp,
University of Cologne, Germany

Reviewed by:
Mariapaola D'Imperio,
Aix-Marseille Université, France
Gorka Elordieta,
University of the Basque Country, Spain
Petra B. Schumacher,
University of Cologne, Germany

Correspondence:
Bettina Braun
bettina.braun@uni-konstanz.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *30 November 2018*
Accepted: *14 August 2019*
Published: *24 September 2019*

Citation:
Braun B and Biezma M (2019) Prenuclear L+H Activates Alternatives for the Accented Word. Front. Psychol. 10:1993. doi: 10.3389/fpsyg.2019.01993*

INTRODUCTION

In intonation languages, utterances may be produced with a series of pitch accents, i.e., tonal targets or movements that are associated with the stressed syllables of accented words, see Example (1) – stressed syllables are underlined.

(1) [We will have to discuss the paper.]$_{IP}$

The utterance in (1) is produced as one intonation phrase (IP), i.e., without further phrasing. The last accent in an IP (or intermediate phrase, in languages that assume two layers of intonational phrases in the prosodic hierarchy, cf. Pierrehumbert, 1980; Grice et al., 2005) is called the nuclear accent. As detailed below, the nuclear accent has received particular attention in the prosodic, semantic, and processing literature. Particularly relevant for this paper is the finding that nuclear accents with certain pitch accents make alternatives more accessible (Weber et al., 2006; Ito and Speer, 2008; Braun and Tagliapietra, 2010; Husband and Ferreira, 2012, 2016; Gotzner et al., 2013; Gotzner, 2014). That is, listeners think of concepts that are contrastively related to the word bearing certain types of nuclear accent (see below), which results in priming effects and more fixations to contrastively related words or objects. Within the semantics/pragmatics literature, it is argued that nuclear accents determine the information structural category of a constituent as focus (shorthand "F"), where a focus constituent is a constituent that evokes alternatives relevant for interpretation (Rooth, 1985, 1992; Krifka, 2008).[1] In this study, we deal with prenuclear accents that signal a contrastive topic interpretation as in the German example in (2) and test whether these accents activate alternatives and if so, whether they do so in the same way as nuclear accents do [unless otherwise indicated, the label *contrastive topic* and the shorthand "CT" is used to refer, descriptively, to constituents with a special prosody that forces a particular interpretation, spelled out by the optional follow up between parenthesis in (2) and whose prosodic features are not of our concern here].

(2) <Die Jungen>$_{CT}$ spielten <Hockey>$_{F}$
the boys played hockey
"The boys played hockey, (but I don't know what the girls did)."

In short (see below for a more detailed discussion on the semantics and interpretation of contrastive topic utterances within the semantics and pragmatics literature), this special prosody (CT-prosody) indicates that the speaker decided to first say something only about a subset of the salient domain, e.g., only about *the boys* in (2), while there are other (contrastive) entities that s/he is not saying anything about.[2] One question this paper tries to shed light on is the status of CT within information structure, i.e., whether this prosody identifies constituents as belonging to a basic notion of information structure (CTs would be then taken to encode a different information category from, e.g., focus), whether it is related to focus, or whether there is no need of an additional information category and focus can also cover these cases. This links the question of CTs as a (possible) notion of information structure with the question about what prosodic cues are used to activate alternatives, in terms of pitch accent type and phonetic realization.

In the remainder of the introduction, we first review the current state-of-the-art on the processing of nuclear vs. prenuclear accents (section "Nuclear vs. Prenuclear Accents"). We then turn to the concept of contrastive topics (section "Theories of Contrastive Topics"); they are interesting because they can be realized with a *prenuclear* accent in German, L*+H, and because contrastive topics are claimed to trigger contrastive alternatives as well. In section "Intonational Realization of Contrastive Topics," the prosodic realization of contrastive topics is reviewed, first for English, then for the target language German. It is shown that the contrast between contrastive and non-contrastive topics in German is realized on a continuum between L*+H and L+H*, with more acustically salient prosodic characteristics in contrastive than non-contrastive contexts, but that German listeners prefer prenuclear L*+H in contexts that trigger a contrastive topic reading. In section "Outline and Hypotheses" we put forth the hypotheses regarding the activation of alternatives.

We then present two visual-world eye-tracking paradigm studies (Cooper, 1974; Eberhard et al., 1995; Tanenhaus et al., 1995), one with naturally produced contours – Experiment 1 – and one with resynthesized contours – Experiment 2 – to investigate four research questions, (a) whether subject constituents that are prosodically marked with prenuclear L*+H lead to more fixations to a contrastive alternative than those marked with prenuclear L+H*, (b) whether the fixation differences occur immediately while the constituent is processed and can hence be attributed to the pitch accent realization, (c) whether there is a difference in fixation pattern between contrastive topic and focus constituents, and (d) whether the activation of alternatives is caused more by pitch accent type or by its phonetic realization (in particular peak height and F0-excursion, which are related to perceived prominence). The answers to these questions will further our understanding on the role of prenuclear accents during speech comprehension, will allow us to contribute to the discussion regarding how to best formally model contrastive topics and overall to the discussion of the taxonomy of information structural categories, and to clarify the role of phonology and phonetic implementation in the activation of alternatives.

Nuclear vs. Prenuclear Accents

The terms nuclear and prenuclear accent stem from the British School (e.g., Halliday, 1967a; Crystal, 1969;

[1]There are different notions of focus in the literature. The different notions of focus range from being the assertive part of an utterance (Lambrecht, 1994), the information that is new relative to the discourse (Firbas, 1975; Halliday, 1985; Vallduví and Engdahl, 1996), to being the constituent evoking alternatives (Rooth, 1992; Steedman, 2000; Krifka, 2008). In recent works, the information structure of utterances is typically established by showing congruence in specific question-answer pair contexts (e.g., Büring, 2009). These notions are not antithetic and can be subsumed under a common core, namely that a focus element is an element that evokes alternatives relevant for the interpretation (see Krifka, 2008 for discussion).

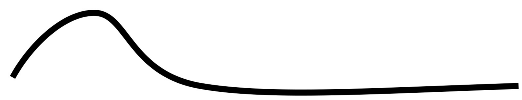
[2]The optional follow up in (2) may also have the same prosodic features as the preceding conjunct, but it does not have to. If spelled out, given that the speaker is clearly dividing the set of entities between boys and girls and we already have the information about the boys, we do not need to use prosody to bring about the contrastive interpretation (e.g., "with respect to the girls specifically, I do not know what they did while about the boys I may know"). We will go in further detail below regarding the interpretation. What needs to be clear is that even though there are different ways to arrive at a contrastive-topic-like interpretation, the question we are focusing here is on how prosody enforces it and on how it can be formally modeled.

O'Connor and Arnold, 1973). In the nowadays dominant framework of autosegmental-metrical phonology (Pierrehumbert, 1980), all pitch accents have the same status. The difference between nuclear and prenuclear accent lies in their distribution in the utterance: nuclear accents form the head of the prosodic phrase and typically occur before a phrase break (intermediate phrase break in case there is one in the intonational phonology of the language, else intonation phase break), i.e., they are the last accent in the phrase. Prenuclear accents precede nuclear accents in the same phrase. In Example (3), if produced as a single phrase, there are hence two prenuclear accents (H* and L+H*) and one nuclear accent (L+H*), followed by a low boundary tone (L−L%).

(3) [We will have to discuss the paper.]$_{IP}$
H* L+H* L+H* L−L%

Nuclear accents have a number of interesting properties. First, they are more prominent to the listener than prenuclear accents, possibly owing to their special structural position. It has been shown that, if a prenuclear and a nuclear accent in the same phrase have the same F0-excursion, the nuclear accent sounds more prominent than the prenuclear one (e.g., Terken, 1991; Gussenhoven et al., 1997), see also Baumann and Winter (2018) for more recent evidence on German. Conversely, a nuclear accent needs less F0-excursion to be perceived as equally prominent as a prenuclear accent. Second, nuclear accents can signal focus and focal information is memorized better (Birch and Garnsey, 1995). Third, in terms of meaning contribution, the choice of nuclear accent type is claimed to signal differences in information status, i.e., whether a referent is new or accessible (e.g., Kohler, 1991; Baumann, 2006), focus location and domain (Eady and Cooper, 1986; Eady et al., 1986; Birch and Clifton, 1995; Baumann et al., 2006; Breen et al., 2010), illocution type (Braun et al., 2018b), as well as attitudinal information, such as sarcasm (e.g., Lommel and Michalsky, 2017).

The past approximately 20 years have accumulated knowledge on how nuclear pitch accents are processed online as the utterance unfolds over time (Dahan et al., 2002; Weber et al., 2006; Chen et al., 2007; Ito and Speer, 2008; Watson et al., 2008; Dennison and Schafter, 2010; Esteve-Gibert et al., 2016; Husband and Ferreira, 2016). In a frequently cited study, Dahan et al. (2002) investigated the effect of accentuation on reference resolution using the visual world eye tracking paradigm. Participants heard two instructions: In the first instruction, they were asked to move an object in a display (e.g., the *candle* in *Put the candle above the triangle*); according to a second instruction they had to move either the same object again (*candle*) or a lexical cohort competitor (*candy*). Object and competitor were either accented (nuclear H* or L+H*, which was not controlled) or unaccented, resulting in four conditions. The results showed that before the cohort competitors were disambiguated segmentally, participants fixated the competitor *candy* more when the noun was accented, suggesting that listeners immediately exploited the relation between pitch accents and discourse structure for reference resolution. Notice that in Dahan et al. (2002) the experimental contrast was between a nuclear accent vs. no accent at all, which is a very prominent intonational contrast. Later studies have also shown that listeners are sensitive to smaller accentual contrasts, i.e., those between different *types* of nuclear accents (Chen et al., 2007; Watson et al., 2008). Moving from discourse effects to the immediate processing of pitch accents, Braun et al. (2018a) recently used the visual-world eye-tracking paradigm to test whether pitch accent type directly affects the fixation of contrastive alternatives, without an explicit context. In Experiment 1a of Braun et al. (2018a), German listeners heard declarative utterances (e.g., *The swimmer wanted to put on flappers*) and watched displays that depicted four printed words: one that was a contrastive alternative to the subject noun (e.g., *diver*), one that was non-contrastively related (e.g., *sports*), the object that had to be clicked (e.g., *flappers*), and an unrelated distractor. That experiment compared a nuclear L+H* accent on the subject [indicating that the subject was in focus, see Example (4)] to a prenuclear L+H* on the subject with a later nuclear accent on the object noun [indicating that the subject was part of a broad focus constituent, see Example (5)].



L+H* L− L%
(4) Der Schwimmer wollte Flossen anziehen.
(nuclear L+H*)
The swimmer wanted flappers wear
"The swimmer wanted to war flappers."

L+H* L+H* L− L%
(5) Der Schwimmer wollte Flossen anziehen.
(prenuclear L+H*)

The results showed that participants directed more fixations to the contrastive alternative when the subject was realized with a nuclear L+H* accent [Example (4)] than when it was realized with prenuclear L+H* accent [Example (5)]. When the utterances were presented with a nuclear H+L* accent on the subject, an accent suitable to mark accessible information (Baumann and Grice, 2006), there was no difference in fixations compared to the prenuclear L+H*. Also, there were no differences in fixations to the visually presented non-contrastive associate (e.g., *sports*). To account for the asymmetric fixation patterns for contrastive and non-contrastive associates, the authors argued against a priming account by which all kinds of related words are more strongly activated when the word is realized with a prominent nuclear accent (L+H*). Instead, they concluded that the fixation data are better captured by the contrast in the semantic/pragmatic import of the two complex accents: the nuclear L+H* accent evokes contrastive alternatives while nuclear H+L* does not. Because there were differential results for the two nuclear accents L+H* and H+L*, such that nuclear L+H* did and nuclear H+L* did not activate alternatives compared to prenuclear L+H*, the authors argued that the fixation differences cannot be due to the status of the accents alone (nuclear vs. prenuclear), but to their interpretations.

Let us briefly discuss an alternative interpretation for the findings in Braun et al. (2018a), which will be addressed in more detail in this paper: the role of perceived prominence. According to e.g., Mixdorff and Widera (2001), accents with a higher peak are judged as more prominent in German (cf. Ladd and Morton, 1997 for English); this effect may not be due to peak height alone, but due to the increased F0-excursion of the tonal movement, as Gussenhoven (2002) pointed out: "[m]any perception experiments [...] have shown that higher pitch peaks sound more prominent, everything else being equal. Interestingly, the effect is not simply due to peak height. Rather, it is an estimate of how wide the pitch excursion is, given some choice of pitch register, and the listener's impression therefore results from an estimate of the pitch span in relation to some choice of pitch register" (Gussenhoven, 2002, p. 50). In the materials of Braun et al. (2018a), the nuclear accents L+H* and H+L* both had a higher peak and a larger F0-excursion than the prenuclear L+H* in the control condition: on average 9 semitones (st) for nuclear accents vs. 5st for the prenuclear accent. So pure peak height or F0-excursion cannot explain the fixation data in Braun et al. (2018a) either. However, we also know that pitch accent type matters for perceived prominence: Baumann and Röhr (2015) tested the prominence of a range of nuclear accent types that followed a prenuclear H* accent. Their findings showed that L+H* (with a F0-excursion of 5st) was judged most prominent, followed in prominence by L*+H (also 5st) and H* (1.2st), all with ratings above 70 on a scale from 0 to 100 (from least to most prominent). H+L* (with an F0-excursion of 6st), the accent that did not result in fixation differences compared to prenuclear L+H* in Braun et al. (2018a), was judged to be less prominent (average prominence rating: 58), despite of its larger F0-excursion compared to nuclear L*+H and L+H* accents. Prenuclear accents were not included in the prominence study by Baumann and Röhr (2015). In a more recent experiment, Baumann and Winter (2018) used the rapid prosody transcription task (Cole et al., 2010) and tested more varied sentence materials and also prenuclear accents. Their data showed that prenuclear accents were less often judged prominent than nuclear accents, but accent type and position (prenuclear/nuclear) were not orthogonally varied so it is not clear whether there is an interaction between the two factors. The perceived prominence of an accent may hence contribute to the activation of alternatives. This is in line with Calhoun (2009) who argued that the more phonetically prominent an accent the more likely a contrastive interpretation. We address the issue of prominence in the activation of alternatives in Experiment 2.

Prenuclear accents have generally been somewhat neglected in the semantic and processing literature, except for studies on their phonetic realization (e.g., Arvaniti et al., 1998; Atterer and Ladd, 2004). Semantically, prenuclear accents have been described as ornamental (Büring, 2007), serving a mostly rhythmic purpose (Calhoun, 2010; Chodroff and Cole, 2018). In a learning paradigm, Kapatsinski et al. (2017) showed that listeners focus more on the nuclear contour and largely ignore the prenuclear accents (cf. Roettger and Cole, 2018 for higher accuracy for whole contours and nuclear tunes compared to prenuclear accents in an artificial language paradigm). Prenuclear L*+H accents may be an exception as this accent type is very prominent as a nuclear accent (Baumann and Röhr, 2015) and its inherent prominence may be used to trigger a CT-reading. This is the accent of interest in the present study.

Theories of Contrastive Topics

There are different theories on how CTs are formalized. While (6) illustrates what is identified as *contrastive topic* constructions in the literature (which assumes a specific prosody that will be reviewed below for English and German), researchers differ on what they take contrastive topics to be and how they are interpreted. We overview the differences between the alternative approaches in (6a–c) below.

(6) Context question: Was haben die Kinder gespielt?
(What did the kids play?)
a. Answer: [[Die [Jungen]$_{\text{Focus}}$]$_{\text{Topic}}$ [haben [Hockey]$_{\text{FOCUS}}$] gespielt. (focus within topic)
b. Answer: [[Die Jungen]$_{\text{CT}}$ [haben [Hockey]$_{\text{FOCUS}}$] gespielt.] (Büring, 2003, 2016)
c. Answer [[Die Jungen]$_{\text{Focus+computation}}$ [haben [Hockey]$_{\text{FOCUS}}$] gespielt.] (Constant, 2014)
"[The boys] [played [hockey]$_{\text{FOCUS}}$]"

All researchers agree that the interpretation of the answers in (6), with the special prosodic features discussed below, can be paraphrased along the lines of "as for the boys, they played hockey" (following Jackendoff, 1972). In these utterances, *the boys* is what is called the contrastive topic constituent while *hockey* is the sentence's (narrow) focus. However, researchers disagree on how we arrive at such an interpretation and on how many basic notions of information structure are necessary to model it (ultimately disagreeing on the taxonomy of information structural categories). These differences are what the contrast in (6a–c) tries to represent (we elaborate on these differences below). The results in this paper won't allow us to discard any of the formal approaches to CT-constructions altogether, but they will allow us to critically evaluate different implementations of such approaches and narrow down the possibilities. On this respect, this paper tries to contribute to a discussion regarding how empirical investigations can inform formal and pragmatic modeling of CT-phenomena and narrow down the landscape. The hope is that future work will continue this discussion. We proceed below to evaluate the different formal approaches.

There are roughly two main camps in the formal semantics and pragmatics literature on contrastive topics (see also Constant, 2014 for an overview): those approaches that appeal to an independent notion of topic (syntactically, semantically or pragmatically defined) and that argue that a contrastive topic is a topic that contrasts with other topics (see Molnár, 1998;

Vallduví and Vilkuna, 1998; Steedman, 2000; Krifka, 2008), and those who do not appeal to any independent notion of topic to understand contrastive topics (see, e.g., Gyuris, 2002, 2009; Büring, 2003, 2016; Tomioka, 2010a,b; Constant, 2014). In fact, a related question in the literature is whether CTs are basic notions of information structure or not. The discussion on CTs is part of a larger debate regarding the taxonomy of information structural categories. For some authors (see, e.g., Krifka, 2008) CTs are topic constituents containing focus (focus being a basic notion of information structure while the status of topic not being that clear). For others (see, e.g., Büring, 2003, 2016) CTs are a basic notion of information structure on their own. Finally, there are others (see, e.g., Tomioka, 2010b; Wagner, 2012; Constant, 2014) for whom CTs are just focus constituents. We provide a brief overview of these approaches and how they differ, and we hope that the sketches below can illuminate the discussion of the empirical results presented in this paper and how they contribute to the discussion of how to best formally model CTs. For the sake of concreteness, we focus below on Krifka (2008) as a representative of theories appealing to independent notions of topic to understand CTs, (6a). We dub this the *focus within topic* approach. We then sketch Büring (2003, 2016) and Constant (2014) as proposals in which understanding CTs does not require an additional notion of topic. These two proposals crucially differ on considerations regarding whether the taxonomy of information structural categories needs to contain both CT and focus (Büring, 2003, 2016), (6b), or whether the notion of focus is enough (Constant, 2014), (6c). We identify these last two approaches by the name of their respective proponents.

Let us start the discussion with the *focus within topic* approach as spelled out in Krifka (2008). Contrastive topics in Krifka (2008) are taken to be cases of *aboutness topics* containing an element marked as *focus*. In this approach to CTs we need both a notion of topic independently defined and a notion of focus. In Krifka's view, the topic constituent is the constituent in the sentence identifying the entity or set of entities under which the information expressed should be stored in the common ground (understood in Stalnakerian terms as the information accepted by participants for the purpose of the conversation). This notion of topic is the notion of *aboutness topic* in Strawson (1964), Halliday (1967b), Reinhart (1981), Gundel (1988), Klein (2008) and goes together with a "structured" view of information update: when accepting the information communicated in an utterance we store it with respect to the topic entity, i.e., we identify the constituent in the utterance that is encoding what the utterance is about, the topic, and the constituent that is encoding what is being said about such entity, the comment, and store that for the given topic the comment has been predicated (this is, e.g., equivalent to the "link" in Vallduví and Engdahl, 1996). In the example in (6a), this would amount to identifying *the kids* as the topic and being able to organize information storage in such a way that we can store a bulk of information specifically about the kids. In particular, in (6a) we are asked to add the information that they played hockey. As for focus, in Krifka's approach a focus element (where focus is a basic notion of information structure) is an element that evokes alternatives relevant for the interpretation [very much the proposal put forward in Rooth (1985), which is also the notion of focus in Büring (2003, 2016) and Constant (2014)]. CTs are then a combination of aboutness topic and focus. In the case of CTs the alternatives that are evoked are alternative topics, i.e., CTs are topics that contrast with other topics (Krifka, 2008, p. 45). Summing up, CT-interpretations are then arrived at by identifying a constituent as being the utterance's aboutness topic and factoring in that it contains focus. This is what we will call the *focus within topic* account. In terms of processing, this view of contrastive topic is compatible with two formal implementations reflecting two processing procedures. One possibility is that conventional linguistic cues (in this case prosodic cues) could both identify a constituent as being the aboutness topic and as containing focus. In this approach the interpretation of the utterance as a contrastive topic would take place online. The other possible implementation involves arriving first at a complete syntactic analysis of the utterance (together with the information-structural analysis) to be able to identify the utterance's aboutness topic and that the focus constituent is indeed within the topic. In this implementation contrastive topics are not processed online.

Let us see how this proposal differs from proposals in which the notion of CT does not depend on an independent notion of topic.[3] Büring (2003, 2016) and Constant (2014) share important features regarding the interpretation of CTs. The interpretation of the sentences in (6b–c), assuming the special prosody discussed below, can be more precisely paraphrased as "as for the boys, they played hockey; the others, I'm not saying (because either I don't know or because I don't want to say)." Büring (and much subsequent work including Constant's) follows the literature on formal discourse models (most importantly Roberts, 1996) and assume that utterances are embedded in a particular discourse structure, where discourse is a hierarchical order of moves organized around (implicit) questions that participants agree on addressing (discourse is a communal inquiry). The assumption in this approach is that "all that is given at the sentential level, conventionally, are certain sorts of presuppositions about the place and function of the utterance in the [intentional structure] of the discourse in which it occurs" (Roberts, 1996, p. 2). Following Rooth (1985), this literature takes focus to be one of the main conventional clues to link the utterance to discourse,[4] since the focus structure of a particular utterance triggers the presupposition that there is a particular question open in the context that is being addressed (i.e., focus anaphora to a contextual question). That this is the case can be illustrated with question-answer pairs. The utterance in (7a) can be the answer to the spelled out question in (7), but (7b)

[3] Büring (2003) takes the stronger position that a notion of *topic* is not necessary in general, while Constant (2014) argues that an independent notion of topic is not necessary to explain CTs but remains agnostic about how necessary it may be to account for other phenomena.

[4] In the Roothian system constituents that generate alternatives relevant for the interpretation are F-marked syntactically; such marking is then reflected in the phonology, although how exactly this last step happens is open to debate. We have nothing to add to the theoretical discussion here. In Rooth's system the focus structure of an utterance links the utterance to discourse by requesting that it be the answer to a question in discourse of a particular form; this is a form of *presupposition* and is cashed formally in Rooth's work by the "∼" operator and its interaction with the focus meaning.

can't. The idea in focus theory is that even when the question is not spelled out, the focus structure allows us to identify what question the speaker is answering: (7a) and (7b) presuppose a different question in the context/discourse [the utterance in (7b) presuppose a question of the form *who drinks coffee?*].

(7) What does Ede drink?
 a. Ede drinks $[\text{coffee}]_F$.
 b. #$[\text{Ede}]_F$ drinks coffee.

In this line of work, the utterance with CT-prosody presuppose a complex question: CT-utterances are analyzed as a partial answer to a (implicit) general complex inquiry of the form, e.g., *who did what?* The responses to the question in (6), assuming the specific prosody, signals that the speaker is resolving only a sub-issue (e.g., *what did the boys do?* in the running example) while s/he is leaving un-answered other contrastive sub-issues (e.g., a contrastive (implicit) question of the form *what did the girls do?*) that should be addressed to provide a complete answer to the complex question.[5] In this way, the speaker is offering only a partial answer to the more general question. Considering that the question that speakers address in the discourse is the topic of conversation, Büring rightfully calls these utterances (as containing) *contrastive topics* (they address a (sub)-topic that contrasts with other topics). What differs between Büring's and Constant's work is how we arrive at this partial-answerhood interpretation. In Büring's system (e.g., Büring, 1997, 2003, 2016), prosody reflects a specific marking in the syntax, CT-marking, see (6b), that comes with its own interpretational rules and lead to the right semantic interpretation (crucially, this marking is different from F(ocus)-marking in the Roothian sense and, hence, CT and focus are taken to be two independent notions of information structure). In Constant's (2014) proposal (see also Tomioka, 2010b; Wagner, 2012), on the other hand, CT-phrases are no more than a F-marked phrase (in the Roothian sense) with special instructions regarding how the evoked alternatives enter into the semantic computation, see (6c).[6] In Constant's system, contrastive topic is not an independent category of information structure. Contrastive topic constituents are just focus constituents (i.e., F-marked constituents in the Roothian sense) plus some instructions regarding how the evoked focus alternatives are to be handled in the interpretation.[7] In this sense, Constant's proposal offers a simpler ontology of information structure categories.

What are the predictions made by these two theories? As said, Büring considers CTs as an information-structural category on their own. This alone may predict a different prosodic realization from F-phrases (the special prosody found in CTs would mark its status as a different information structural category). Notice, however, that in Büring's theory the alternatives evoked by F-marked phrases and CT-marked phrases are different: syntactic F-marking evokes alternative propositions while syntactic CT-marking evokes alternative questions. In Constant's system CTs are focus phrases. This approach hence makes the prediction that CT-phrases evoke alternatives in the same way as F-phrases do. Constant's theory also makes predictions regarding the different prosodic realizations found in CT-phrases and F-phrases by virtue of their syntax. CT-phrases in Constant's system are taken to be in the left periphery, either because they are moved there or because they are generated there, and it is this syntax that is responsible for the special prosody. How do we choose between the two systems? In what follows we sketch our reasoning in this paper.

The empirical investigation presented in this paper is related to how alternatives are activated in CT-constructions in contrast to what we find in narrow focus. That the alternatives that are evoked in CTs are different from those in narrow focus constructions (e.g., alternative propositions vs. alternative questions, as in Büring's system) does not warrant a prediction that we should observe differences in the way alternatives are evoked/activated in contrastive topics constituents and focus constituents but, if we did observe such difference, we may consider it as partial support for contrastive topics being different from focus (against Constant's proposal). At the same time, if there is no difference between how alternatives are evoked in contrastive topics constituents and focus constituents, we would lack support for a system that considers contrastive topics different from focus. That is, everything else being equal, if we are to choose between two systems, one simpler than the other, we need arguments to support that the more complex system is justified, e.g., in terms of processing. One way to do that is by showing that the way alternatives are evoked for contrastive topic and focus is different, explaining why we need two different information structural categories (cashed out formally in a different syntactic marking and interpretational mechanisms). If two models can derive the same results, in the

[5]This more precise paraphrase is not applicable to *focus within topic* proposals of CTs. In Büring's and Constant's approach the formal system leads to the prediction that participants are (collectively) committed to also address the other sub-questions (i.e., to provide a complete answer for the more general complex question). In Krifka's version of the *focus within topic* approach this is not encoded (although it can be derived as an inference).

[6]Constant's (2014) proposal falls into what he calls *configurational* accounts of contrastive topics, i.e., proposals that take a CT-phrase to be simply an F-marked constituent in a particular configuration. Other proposals of this sort include Tomioka (2010b) and Wagner (2012). The crucial difference between Constant's (2014) proposal and other configurational accounts lies in the range of data the different proposals can explain within the characterization of CT (e.g., Wagner, 2012) argues that the configuration in which the constituent with a contrastive focus marking precedes the exhaustive focus is explained with different means from the configuration in which the contrastive topic constituent follows the exhaustive focus), as well as the predictions made regarding the phonology-syntax interface (i.e., Constant's proposal can derive differences in the prosody of contrastive topic phrases vs. that of exhaustive focus, while those prosodic differences are not that easy to derive in other configurational proposals). The reader is referred to Constant (2014) for detailed discussion of the differences between different configurational proposals.

[7]In Constant's system the CT-phrase is either (covertly or overtly) moved to the left periphery or base generated there. The CT-phrase is an F-marked phrase that composes with the rest of the sentence via a CT-operator delivering the right interpretation within the (Roothian) focus dimension. This operator leaves the ordinary meaning intact. The desired interpretation of utterances with CT-marking is arrived at via old (Roothian) focus anaphora to a contextual question. In Büring's system, we require a new dimension of meaning: besides the old focus meaning in Rooth's system, Büring makes use of the CT-value, which requires its own compositional rules. Büring (2016) is explicit in that CT-interpretations are the result of a *conventional implicature* encoded in the CT-marking.

absence of support for a more complex model we shall prefer the simpler approach.

Regarding how alternatives are evoked in CTs we investigate whether, as in the case of focus, alternatives are evoked online. Both (6b) and (6c) are compatible with alternatives being evoked online. However, for (6a) we saw that there are different possible implementations. The analysis is compatible with alternatives being evoked as soon as the accent is processed (online processing), but it is also compatible with late activation, once the listener has already assigned a syntactic analysis of the constituent as topic.

All proposals depicted in (6) predict that there is a difference between the answers in (6) and (8). Given the provided context-question, an exhaustive (neutral) answer[8] is not expected to have the same prosodic marking as the CT-utterance in (6).

(8) Context question: Was haben die Jungen gespielt?
(engl. What did the boys play?)
Answer: Die Jungen haben Hockey gespielt.
"The boys played hockey."

An important question addressed in this paper concerns the way we process utterances triggering CT-interpretations and whether this differs from the processing of focused constituents.

Intonational Realization of Contrastive Topics

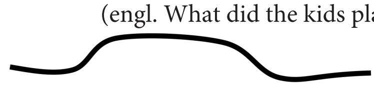
Since utterances with contrastive topic and focus constituents and broad focus utterances can have the same (surface) structure [see Examples (6) and (8)], when heard out of context, it is the intonational realization that distinguishes the interpretation of the grammatical subject as contrastive topic or focus or neither. Contrastive topics are often realized with different pitch accents from focal constituents. In English, Jackendoff (1972) described the prosodic realization of contrastive topics in English as B-accents (falling-rising contours) and foci as A-accents (falling contours), see Example (9). In the autosegmental-metrical framework, the B-accent contour is a complex phenomenon, represented as L+H* L−H% (authors also consider L*+H as a possible complex accent for CT-phrases, see, e.g., Constant, 2014), while the A-accent contour is equivalent to H* L−L%, the prosodic realization of an exhaustive focus in English.

(9) Question: What about Fred? What did he eat?

Answer: [(Fred)$_{ip}$]$_{IP}$ [(ate the beans.)$_{ip}$]$_{IP}$
B-accent A-accent
(Jackendoff, 1972)
L+H* L−H% H* L−L%
(Pierrehumbert, 1980; Büring, 2003; a.o.)

In German, however, contrastive topics are realized with a prenuclear rising L*+H accent, while the (exhaustive) focus is realized as falling nuclear accent (Féry, 1993, p. 131). Unlike in English, there is typically no IP break between the contrastive topic and focus constituent (and hence there is no L−H% boundary tone). In German, the contrastive topic and the focus constituent are often produced in the same prosodic phrase. It is also often argued that the F0 contour between the rising accent on the contrastive topic and the fall on the focus remains high, resulting in the so-called hat pattern (originally described for Dutch by Cohen and 't Hart, 1967)[9]. This realization is exemplified in Example (10), using the prosodic notation of the GToBI, German Tone and Break Indices, system (Baumann et al., 2001; Grice et al., 2005). German hence marks contrastive topics with a prenuclear accent. The prenuclear accent is prototypically an L*+H, an accent that is judged as one of the two most prominent accents when placed in nuclear position (Baumann and Röhr, 2015). The nuclear accent on the focus constituent, H*, is one that is not judged very prominent.

(10) Context question: Was haben die Kinder gespielt?
(engl. What did the kids play?)



Answer: [(<u>Die Jun</u>gen haben <u>Hock</u>ey gespielt.)$_{ip}$]$_{IP}$
(lit. The boys have hockey played)
L*+H H* L−%
GToBI notation

Experimental studies with identical sentences in different information structures showed that the prosodic difference between utterances identified as triggering a CT-reading [Example (10)] and those lacking this interpretation is not categorical (Braun, 2005, 2006, 2007). Instead, contrastive topics are typically realized with a later and higher peak and longer duration than the prenuclear rise in utterances without CT-interpretations. The hat pattern is not mandatory either. From the listeners' perspective, while the prosodic contrast in the prenuclear accent in CT- and non-CT utterances is not necessarily categorical, prenuclear L*+H is interpreted as contrastive topic, prenuclear L+H* is not. This was shown in a binary forced-choice context-matching experiment, in which participants received a written context (e.g., '*Jetzt geht es um einen Sohn und eine Tochter. Der Sohn beschäftigt sich mit Latein und...*' "The next story is about a son and a daughter. The son is occupying himself with Latin and...") and heard a target sentence (*Die Tochter beschäftigt sich mit Mathe.* "The daughter is occupying herself with mathematics.") in one of eight conditions, manipulating prenuclear accent type (L*+H vs. L+H*), nuclear accent type (H* vs. H+L*) and the F0-transition between prenuclear and nuclear accent (high plateau vs. dip). The highest acceptance came from utterances with a prenuclear L*+H accent and a nuclear H+L* accent, while the F0 transition

[8]Exhaustivity can be derived either from the pragmatics of question-answer pairs or from a particular semantics for focus. Notice, however, that exhaustivity is not a feature of the Roothian focus-semantics.

[9]In the German literature, this contour is known under the names *Hutkontur* "hat pattern" (Mehlhorn, 2001; Steube, 2001) "bridge contour" (Wunderlich, 1991; Büring, 1997), or *Wurzelkontur* "root contour" (Jacobs, 1982).

between the two did not matter (81.6% for the high plateau, 89.3% for the dip). It is interesting to note that the preferred focus accent in CT-constructions in German is nuclear H+L*, an accent type that is not judged particularly prominent in Baumann and Röhr (2015). In a context that did not trigger a CT-interpretation (CONTEXT: Die Tochter beschäftigt sich mit Mathe. "The daughter is occupying herself with mathematics.", TARGET., weil sie morgen eine Klausur schreibt. "... because she will have a test tomorrow."), participants gave highest agreement to contours with a prenuclear L+H* accent on the subject and a nuclear H* accent, irrespective of the F0 transition (69% for the high plateau, 68% for the dip). Given all these results, we will use a prenuclear L*+H accent on the subject constituent and a nuclear H+L* accent as focus for the CT-condition in the experiments reported below. Since the F0 transition between the prenuclear and nuclear accent did not have an influence on perception, we stuck to one pattern, the hat pattern, which was more natural for the speaker. Regarding the *phonetic* implementation of prenuclear accents, offline acceptability studies have shown that participants find prenuclear rising accents with higher peaks more appropriate in contexts that triggered a CT-interpretation, accents with later but lower peaks were less acceptable but more appropriate than rises with earlier and lower peaks. In unmarked all-new contexts (Braun, 2004, 2005), there was no preference. Note that prenuclear L*+H has also been reported as neutral prenuclear accent in Truckenbrodt (2002), who analyzed a not further specified sample of Southern German and Austrian speakers.

Outline and Hypotheses

While the interpretation of the CT-constituent is often linked to contrast and some theories even link the CT-constituent directly to focus (see discussion above) this has not been supported by empirical findings in the literature yet. If CT-constituents were shown to activate alternatives, this would be the first demonstration that CT is processed like focus and that certain types of prenuclear accents (in addition to nuclear accents) have the potential to do so. Furthermore, depending on how this activation compares to the activation of alternatives found for utterances with narrow focus, the findings could provide empirical support to theories linking CT to focus in its treatment.

We use the visual-world eye-tracking paradigm with printed words (McQueen and Viebahn, 2007), which allows us to study the processing of contrastive alternatives without interference from visual relatedness (Huettig and McQueen, 2007). For the sake of comparability, we closely replicate Experiment 1a in Braun et al. (2018a), see examples (4) and (5). In Experiment 1 in this paper we compare two intonation conditions, naturally produced prenuclear L*+H (contrastive topic, CT, condition) to naturally produced prenuclear L+H* (broad focus control condition). We measure participants' fixations toward these referents while they process utterances in the two intonation conditions. A higher number of fixations to the contrastive associate in the contrastive topic compared to the control condition is interpreted as increased activation of the contrastive alternative in the contrastive topic condition. Note that the term "activation" is understood here as shorthand for "consider as lexical or conceptual alternatives," In Experiment 2, we manipulate the intonation contours (PSOLA resynthesis) to reduce phonetic differences between contours.

Based on the semantic literature and the available processing data, we pose the following hypotheses on the activation of alternatives. The literature reviewed above results in a number of conflicting hypotheses on the role of prenuclear accents in processing (H1), on the comparison of contrastive topics accents and focus accents (H2), and on the role of F0-excursion of an accent for the activation of alternatives (H3). In what follows, we briefly lay out the possible hypotheses and advance some possible points of contention working against them.

H1. The available processing literature suggests that prenuclear accents are not processed as deeply (semantically) as nuclear accents. From that perspective, one would expect no differences in fixations between prenuclear L*+H and prenuclear L+H*. However, since prenuclear L*+H has the potential to signal CT-constituents (among other things), we predict that prenuclear L*+H leads to more fixations to the contrastive associate than prenuclear L+H* accents.

H2. Given that prenuclear L*+H leads to a CT-reading, according to semantic/pragmatic proposals we predict that this accent has the same potential to activate contrastive alternatives than the nuclear L+H* focus accent of Experiment 1a in Braun et al. (2018a). If CT equals focus, we expect a similar effect size and a similar timing as for the focus data of Experiment 1a in Braun et al. (2018a).

H3. If a large F0-excursion is the decisive factor for the activation of alternatives, we predict that the fixation difference disappears when using resynthesized stimuli with the same F0-excursion of the rise for prenuclear L*+H and L+H*. These two accents did not differ in perceived prominence in Baumann and Röhr (2015) in nuclear position, where they had the same F0-excursion. If the interpretation of the accent type that is relevant, we hypothesize the same fixation differences between prenuclear L*+H and prenuclear L+H* with resynthesized stimuli.

Hypotheses H1 and H2 are tested in Experiment 1, hypothesis H3 mainly in Experiment 2. Note that the experimental results with respect to H1 and H2 will allow us to discuss the different semantic/pragmatic formal theories in view of the psychological reality of contrastive topics.

EXPERIMENT 1

Methods

Participants

Forty native speakers of German between 19 and 33 years (average 25.7 years) participated for a small fee. Twenty-eight were female, 12 male. They were unaware of the purpose of the experiment and had not taken part in experiments involving similar materials. All participants reported to have normal hearing and had normal or corrected-to-normal vision. Written informed consent was obtained.

Materials

Sentences and visual displays

The experiment used the same sentence materials and displays as in Braun et al. (2018a). There were 24 experimental sentences and 24 filler sentences. All experimental sentences started with a subject-NP (see **Table A1** in the **Appendix**), followed by a disyllabic auxiliary (*wollte* "wanted to", *hatte* "had", *konnte* "could", and *sollte* "should"), an object noun and a non-finite verb (*Der Turner hatte Blasen bekommen* "The gymnast had gotten blisters"). Most of the subject-referents had penultimate stress and between two and four syllables. None of them had ultimate stress. The filler sentences were similar to the experimental sentences and also started with a definite subject-NP followed by a disyllabic auxiliary. However, they occassionally contained disyllabic verbs and temporal adverbials.

The words for the display in experimental trials had been selected as follows. For each of the subject nouns, there was one noun that was contrastively related and one that was non-contrastively related. The non-contrastive associate was collected in a free association task. Participants saw one noun at a time (e.g., gymnast), printed on screen, and had to type in the first word that came to their mind (e.g., sports). Due to this procedure of collecting highly active non-contrastive associates, these associates do not all have the same relation to the auditory target, i.e., some stand in a hyponym-hyperonym relation, others in a part-whole relation or refer to a typical instrument or location. While the hyponyms and hypernyms would qualify as replacements for the auditory target, the part-whole relations do not. It was not possible, however, to find enough non-contrastive associates with the same relation to the target. To collect the contrastive associate, participants saw a sentence fragment with a negated subject noun (e.g., "Not the gymnast had gotten blisters but the...") and had to type in the most plausible continuation. For both the contrastive and the non-contrastive associates we chose the most frequent responses making sure that they differed from each other, were not onset competitors and had similar word lengths and lexical frequencies (factors that are known to affect fixation behavior, cf. Dahan et al., 2001; Kliegl et al., 2004). The average association strength, lexical frequency and number of characters of the selected contrastive and non-contrastive associates were matched, see **Table 1**. Each experimental trial showed the contrastive and non-contrastive associate, the grammatical object that had to be clicked as well as an unrelated distractor. The four words in any given experimental trial differed in onset letters.

In filler trials, the display showed the contrastive associate, the grammatical object that had to be clicked, a word that was non-contrastively related to the object and an unrelated distractor. In filler trials, the four words also differed in onset letters.

Recordings

The control condition (see **Figure 1**) and the fillers were the same as in Braun et al. (2018a). The experimental utterances (CT condition) were recorded anew, by the same female speaker of German under the same conditions (44.1 kHz, 16 Bit), see **Figure 2**. All sentences in the experiment were preceded by the prelude *Und ich habe gehört* "And I have heard," to increase the preview time for the words in a natural way. This prelude was recorded once and spliced in front of all sentences with a pause of 1000 ms in-between.

Acoustically, prenuclear L*+H (contrastive topics) differed from prenuclear L+H* in that they had a significantly later alignment of the L and H targets, a larger F0-excursion, and a longer duration of the stressed syllable, of the F0-rise and the entire subject-NP compared to prenuclear L+H*. The mean values and standard deviations for each of these measurements in the two intonation conditions are listed in **Table A2** in the **Appendix**. The sound files are availabe at **Supplementary Data Sheets S1–S3**.

Procedure

Intonation condition was manipulated as a within-subjects factor (but for every participant between-items), i.e., each participant

TABLE 1 | Average association strength, lexical frequency and number of characters (and standard deviations) of contrastive and non-contrastive associates to the subject nouns.

	Contrastive associate	Non-contrastive associate
Association strength (percentage)	30.3 (*SD* = 14.9)	27.9 (*SD* = 16.6)
Lexical frequency (occurrences per million)	1.5 (*SD* = 2.1)	4.6 (*SD* = 5.5)
Number of characters	6.8 (*SD* = 1.5)	5.9 (*SD* = 1.9)

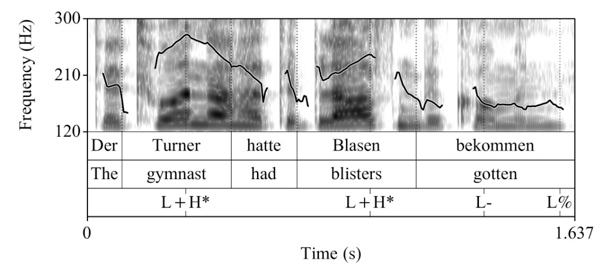


FIGURE 1 | Example realization of a sentence recorded in the broad focus control condition (prenuclear L+H*).

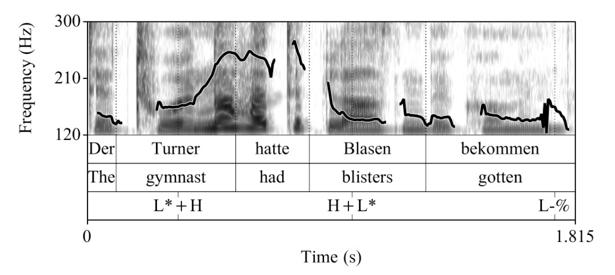


FIGURE 2 | Example realization of a sentence recorded in the contrastive topic condition (prenuclear L*+H).

saw all of the 24 experimental trials, but each target sentence was presented in only one of the two intonation conditions (totaling in 12 trials for each intonation condition). Across the experiment, the position of each of the different types of printed words was balanced (i.e., it occurred equally often in the upper left and right, lower left and right parts of the screen).

Two basic experimental lists were constructed, following a Latin Square Design. Each list further contained all the filler sentences. The two basic experimental lists were pseudo-randomized four times with the restriction of at most three experimental trials in a row (but at most two of the same intonation condition). After each block of five trials, an automatic drift correction was initiated. In total, we had eight experimental lists, to which participants were randomly assigned (five participants for each list).

Every trial started with a fixation cross which was shown until participants clicked on it. In all trials, the same token of the prelude (with a duration of 897 ms) was used. This was followed by a 1000 ms silence, after which the target utterance was auditorily presented. After participants had clicked on the respective object, there was a 1000 ms inter-trial interval. Eye-movement data (fixations, blinks, saccades) were recorded throughout the experiment.

The testing procedure was the same as in Braun et al. (2018a). Participants were tested individually in a sound attenuated room at the University of Konstanz. They were instructed in writing to listen to the utterances and to click on the object that is mentioned therein as quickly as possible. The instructions gave an example to make sure that participants knew what the object is.

Participants sat at a distance of approximately 70 cm from a 20 inch LCD screen, so that they could freely move the computer mouse. They rested their chin on the provided chin rest. Their dominant eye was calibrated with an SMI Eyelink 1000 system (pupil and corneal reflection at a sampling rate of 250 Hz). The same sampling rate was used during trials. The auditory stimuli were presented via headphones (Sennheiser PMX90) at a comfortable loudness.

Results

The eye-tracking data were processed as in Braun et al. (2018a). That is, the eye movement record was sampled in 4 ms steps and automatically parsed into saccades, fixations, and blinks by the EyeLink software (using normal saccade sensitivity). Only fixations were further processed. They were automatically coded as pertaining to a given word if they fell within a rectangle of 100 × 100 pixels, centered on the middle of that word. The grand average of evolution of fixations to the four words in the two intonation conditions is shown in **Figure 3** (using the VWPre package in R, see Porretta et al., 2017). The gray vertical dashed lines indicate the segmental reference points, i.e., word boundaries from left to right. Note that it takes approximately 200 ms to launch a saccade (Fischer, 1992; Matin et al., 1993; Altmann and Kamide, 2004), which is also the delay in our studies: The fixations to the target (the grammatical object that had to be clicked, blue line in **Figure 3**) increased at approximately 1000 ms after utterance onset in the broad focus condition, i.e., approximately 200 ms after the onset of the grammatical object. The same delay of 200 ms is observed in the prenuclear L*+H condition and is hence a good approximation for the time it took participants to launch saccades based on the auditory input. Hence, only after this time fixations can be interpreted as a response to the acoustic signal.

The interesting line for our research question is the red line in the time window from about 330 ms to 770 ms (i.e., 200 ms after the onset of the subject noun till 200 ms after its offset). This line shows fixations to the contrastive associate while participants were processing the subject noun. In **Figure 4**, the fixations to the contrastive associate in the two intonation conditions are compared directly.

For statistical analysis we analyzed participants' fixations to the contrastive referent in consecutive 100 ms steps (cf. McQueen and Viebahn, 2007). We calculated the empirical logits of fixations to the contrastive associate in consecutive 100 ms windows starting from 100 ms after the onset of the utterance until 800 ms after its onset, dividing the fixations to that word by fixations that were directed elsewhere. A constant of 0.5 was added to both the denominator and the numerator (Barr et al., 2011). Empirical logits were analyzed using linear mixed effects regression models with *intonation condition* (prenuclear L*+H vs. L+H*) as fixed factor (dummy coded) and random intercepts for *participants* and *items* (Baayen, 2008; Baayen et al., 2008). The model further included random slopes for the two within-group factors when this improved the fit of the model, as determined by LogLikelihood comparisons, using the R-function anova(). *P*-values were calculated using the Satterthwaite approximation of degrees-of-freedom in the R-package lmerTest (Kuznetsova et al., 2016), which is based on lme4 (Bates et al., 2014).

In the time window 500–600 ms after the onset of the utterance, there were significantly more fixations to the contrastive associate in the contrastive topic condition (average logits = −1.7) than in the broad focus control condition (average logits = −2.3, ß = 0.56, $SE = 0.19$, $df = 922$, $t = 2.9$, $p < 0.005$), see **Table 2** for *p*-values in all time windows. Note that there were no other significant differences in fixations to the contrastive associate in the entire time window shown in **Figure 3**. Given the time needed to plan a saccade, this difference is well within the time during which participants were processing the subject noun (170–270 ms after the onset of the subject noun, a period in time when all items are already unique when considering part-of-speech, grammatical gender, segments and stress, as indicated by a CELEX search). Note that fixations to the contrastive associate were numerically higher in the prenuclear L*+H than in the prenuclear L+H* condition from the start of the utterance, but this difference was not significant. At the moment, we don't have an explanation for this slight preference of the contrastive associate in the contrastive topic condition.

In both intonation conditions, there were also many fixations to the non-contrastive associate, but these fixations to the non-contrastive associate were not affected by intonation condition (second row of **Table 2**). There were more fixations to the target (i.e., the grammatical object that had to be clicked) in the broad focus control condition than in the contrastive topic condition.

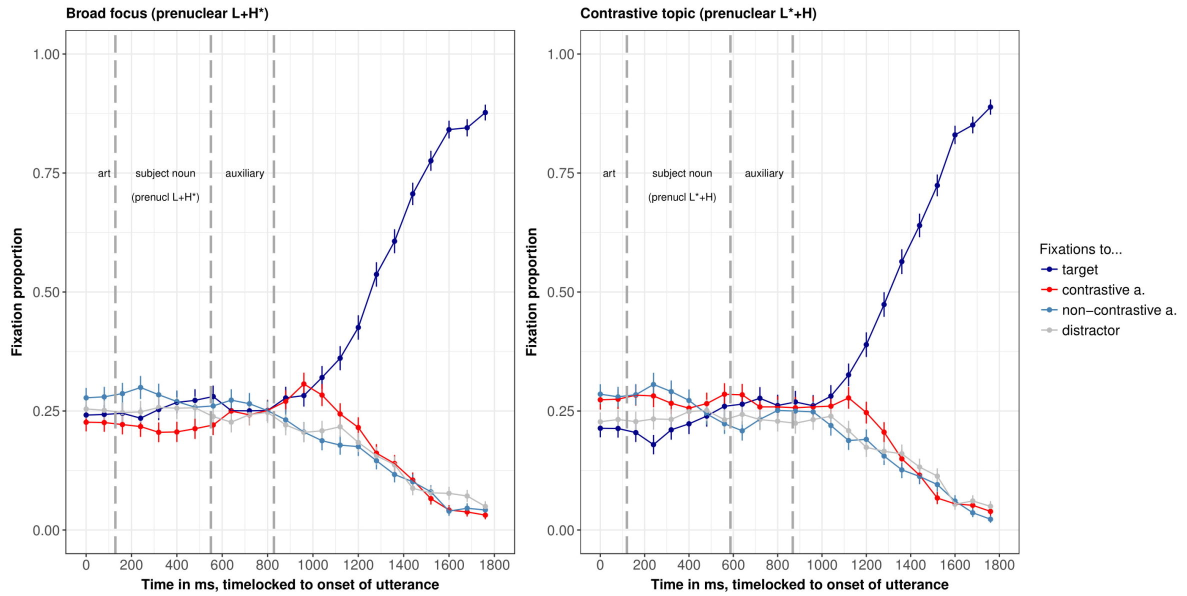


FIGURE 3 | Grand averages of fixation proportions to the four words on screen, split by intonation condition (**left panel:** prenuclear L+H*, **right panel:** prenuclear L*+H), in 80 ms bins of Experiment 1. Whiskers show standard error. The line of interest is the red line, which shows fixations to the contrastive associate.

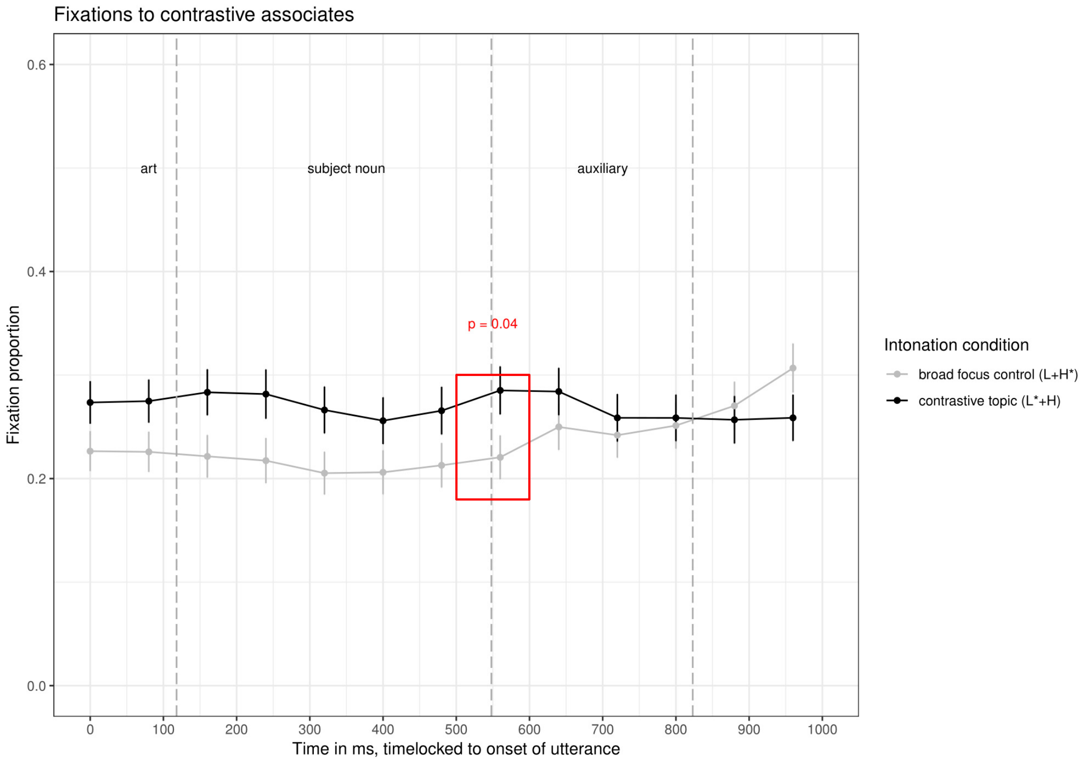


FIGURE 4 | Comparison of fixations to the contrastive associate in the two intonation conditions of Experiment 1.

This effect approached significance in the time windows from 200–500 ms after the onset of the sentence (see **Table A3** in the **Appendix**). This is the opposite pattern as for the fixations to the contrastive associate, which suggests that target fixations are reduced in the contrastive topic condition because of increased fixations to the contrastive associate.

We then compared whether the effect of intonation condition was stronger here, in the prenuclear L*+H condition than in

TABLE 2 | Summary of *p*-values of comparisons of fixations to the contrastive associate (first row) and non-contrastive associate (second row) across intonation conditions in consecutive 100 ms analysis windows of Experiment 1.

	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms
Contrastive associate	$p = 0.1$	$p = 0.1$	$p = 0.07$	$p = 0.1$	$\mathbf{p < 0.005}$	$p < 0.07$	$p = 0.3$
Non-contr. associate	$p = 0.9$	$p = 0.8$	$p = 0.1$	$p = 0.9$	$p = 0.4$	$p = 0.4$	$p = 0.2$

Bold face indicates a significant difference at α = 0.05. The subject noun starts on average 130 ms after the onset of the sentence; it ends on average 580 ms after the onset of the sentence (averages over both intonation conditions).

the nuclear L+H* (contrastive focus) condition of Experiment 1a in Braun et al. (2018a). In that experiment, there was an effect of intonation condition in the same time window, but with a smaller magnitude (ß = 0.4 in Braun et al. (2018a) compared to ß = 0.56 in this experiment). To this end, we combined the data set and calculated the interaction between *experiment* and *condition* (contrastive topic/focus vs. broad focus control). The model showed no interaction between *experiment* and *condition* ($p = 0.5$); there was only a significant effect of *condition* in the combined data set, with more fixations to the contrastive alternative in the contrastive accents (nuclear L+H* and prenuclear L*+H) than in the control condition (ß = 0.6, $SE = 0.19$, $df = 1839.9$, $t = 2.9$, $p = 0.003$). The lack of an interaction does not allow for strong conclusions. An additional Bayes Factor analysis indicated that the simpler model was more than 200 times more likely than the model with the interaction (Morey and Rouder, 2018). This suggests that the activation of contrastive alternatives is not different for nuclear L+H* accents and prenuclear L*+H accents.

Discussion

The eye-tracking data showed that participants fixated more on contrastive associates to the subject constituent when it was produced with a prenuclear L*+H accent compared to a prenuclear L+H* accent. The difference was significant in the time window from 500–600 ms after the onset of the utterance, i.e., immediately while participants were processing the subject noun. We interpret these differences in fixations to the contrastive associate as evidence for an activation of alternatives upon hearing subjects with a prenuclear L*+H accent as compared to prenuclear L+H*. Given the lack of a difference for fixations to the non-contrastive associate, the data speak in favor of a model in which prenuclear L*+H is a contrastive accent in the sense that it leads to an increased activation of contrastive alternatives. Note that this difference in fixations to the contrastive associate for prenuclear L*+H (vs. prenuclear L+H*) is the same as the difference in fixations reported for comparison of nuclear L+H* (contrastive focus vs. prenuclear L+H*) reported in Experiment 1a in Braun et al. (2018a). It is of similar magnitude and occurs at the same time window, specifically between 500–600 ms after the onset of the utterance. The data hence suggest that nuclear L+H* and prenuclear L*+H have the same potential to activate alternatives, vis-à-vis a non-contrastive prenuclear L+H* accent. This finding has interesting implications for the modeling of contrastive topics (see General Discussion).

We now focus on the time course of the effect of intonation condition to determine which part of the contour may have resulted in the activation of alternatives. We observe significant differences in fixations to the contrastive associate in the time window 500–600 ms after utterance onset. These fixations are triggered by acoustic information that occurred around 300–400 ms after utterance onset the latest (170–270 ms after the onset of the subject noun). This suggests that participants' fixations are guided directly by the F0 information before and on the stressed syllable. Ritter and Grice (2015) already showed that German listeners are particularly sensitive to this "onglide" information, but only for nuclear accents. We add to this that prenuclear accents do not differ in this respect. Note that in this analysis window, only information on the pitch-level of the accented syllable is available (L* vs. H*) and some information on the direction (rising or falling), but no information on the following pitch movement (dipping in broad focus, high plateau in contrastive topic condition). It hence seems that the pitch accent alone is sufficient to trigger the contrastive interpretation. This ties in with offline acceptability judgments, in which participants judged utterances with a combination of prenuclear L*+H followed by a nuclear H+L* nuclear accent as more appropriate in a contrast that elicits a CT-interpretation, while the intervening pitch contour (the presence/absence of hat contour) had no effect (Braun and Asano, 2013). It is also consistent with findings on German that suggest that the onglide (the F0-information prior to the stressed syllable) is important for interpretation (Ritter and Grice, 2015).

EXPERIMENT 2

Experiment 2 tested whether the differences in fixations to the contrastive alternatives are solely due to the differences in accent type (prenuclear L*+H vs. L+H* here) or due to the differences in phonetic implementation of these accent types (in particular the peak height and the F0-excursion of the rise and the concomitant differences in perceived prominence). Since there are different opinions on whether prominence is related more to F0-excursion or the scaling of the tonal targets, we manipulated the F0-contour of both intonation conditions to make their F0-excursions (and the scaling of the low and high tonal targets of the accents) the same. Specifically, we (a) raised the low tonal target in the L*+H condition, while keeping the high tonal target unaltered (making the CT accents less prominent under the view that L*-accents are more prominent the lower the L-target and under the view that F0-excursion is related to perceived prominence) and (b) lowered the entire register of the L+H* condition, to have exactly the same F0-scaling for low and high tonal targets and a similar degree of unnaturalness induced by the resynthesis procedure.

Methods

Participants

A different set of 40 speakers of German, recruited from the same subject pool, participated for a small fee. They were aged between 19 and 30 years (average 22.5 years, 32 female, 8 male). The participants were unaware of the purpose of the experiment and had not taken part in experiments involving similar materials. All participants reported to have normal hearing and normal or corrected-to-normal vision. Written informed consent was obtained.

Materials

The sentences and the visual displays were the same as in Experiment 1. All recordings were manipulated to achieve a similar F0-excursion for the contrastive topic and broad focus control stimuli and to achieve a matched sound quality. The recordings of the contrastive topic condition were first stylized [using the stylize pitch (2 semitones) function in praat, cf. Boersma and Weenink (1992-2011)]. Then, the low F0-values prior to the F0-rise were shifted up by 20 or 30 Hz, the choice depending on the naturalness of the resynthesis. Most utterances were shifted up by 30 Hz. The low F0-values after the nuclear accent were shifted up by the same amount. Furthermore, the F0-maximum was shifted up by 10 Hz for four recordings which had very low F0-maxima. The recordings of the control condition were also stylized and uniformly shifted down by 20 Hz, the fillers were stylized and shifted down by 10 Hz (a 20 Hz shift did not result in naturally sounding stimuli, so we sacrificed similarity of resynthesis procedure for naturalness in the case of fillers). This manipulation only changes the register. The acoustic realization of the resynthesized stimuli is shown in **Table A4** in the **Appendix**. Crucially, the stimuli in the contrastive topic condition and the control condition did not differ in the F0-excursion of the pitch rise ($p > 0.9$), in the F0-value of the minimum before the rise ($p > 0.3$) and the F0-value of the maximum ($p > 0.1$).

Example comparisons between the resynthesized F0-contour across Experiments are shown in **Figure 5** for the broad focus control condition and in **Figure 6** for the contrastive topic condition.

Procedure

The experimental lists and the procedure were identical to Experiment 1.

Results

The evolution of fixations to the four words on screen over time is shown in **Figure 7**, the comparison of fixations to the contrastive alternative over time in **Figure 8**.

The results were analyzed in the same way as for Experiment 1. The analysis of fixations in subsequent 100 ms bins showed a significant effect of intonation condition in the time window from 100–200 ms and 700–800 ms (see **Table 3**, first row) after the onset of the utterance. In the 100–200 ms time window,

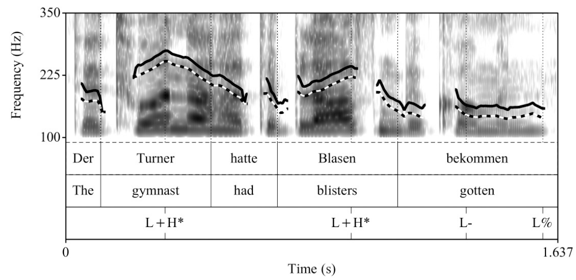


FIGURE 5 | Comparison of F0-contours in the control condition (solid line: original contour of Experiment 1, dotted line: resynthesized contour of Experiment 2).

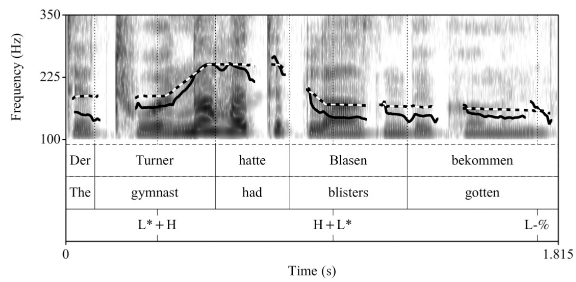


FIGURE 6 | Comparison of F0-contours in the contrastive topic condition (solid line: original contour of Experiment 1, dotted line: resynthesized contour of Experiment 2).

participants' fixations are not yet triggered by acoustic material from the stimulus, so it is difficult to understand the source of these differences. In the 700–800 ms time window, which clearly results from acoustic information in the subject noun, the average logits to the contrastive associate in the prenuclear L*+H condition was −1.56, compared to −2.00 in the control condition (β = 0.43, SE = 0.2, df = 896, t = 2.1, p = 0.03). The time window of significant differences between prenuclear L*+H and prenuclear L+H* is hence 200 ms later than in Experiment 1, while participants were starting to process the auxiliary following the subject. To test whether the differences in analysis windows between Experiment 1 and 2 are statistically significant, we pooled the data of both experiments and tested for an interaction between experiment and intonation condition. The interaction was not significant in any of the analysis windows (500–600 ms: p = 0.2, 600–700 ms: p = 0.8, 700–800 ms: p = 0.4). In all three analysis windows, there was only an effect of intonation condition (500–600 ms: p = 0.004, 600–700 ms: p = 0.02, 700–800 ms: p = 0.03).

Similar to Experiment 1 [and the Experiments in Braun et al. (2018a)], fixations to the non-contrastive associate did not differ across conditions (see **Table A5** in the **Appendix**).

Discussion

The results of Experiment 2 showed that pitch accent type (prenuclear L*+H vs. prenuclear L+H*) mattered for the interpretation and processing of subject constituents. As in

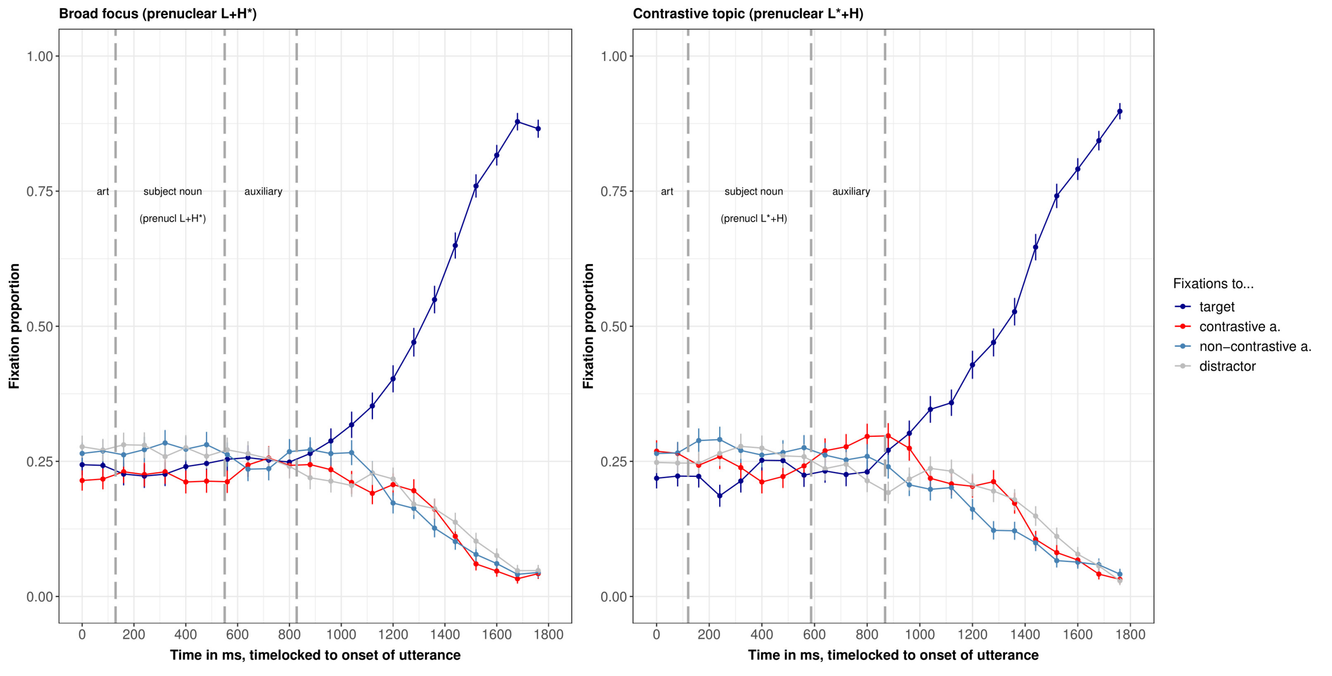


FIGURE 7 | Grand averages of fixation proportions to the four words on screen, split by intonation condition (**left panel:** resynthesized prenuclear L+H*, **right panel:** resynthesized prenuclear L*+H), in 80 ms bins of Experiment 2.

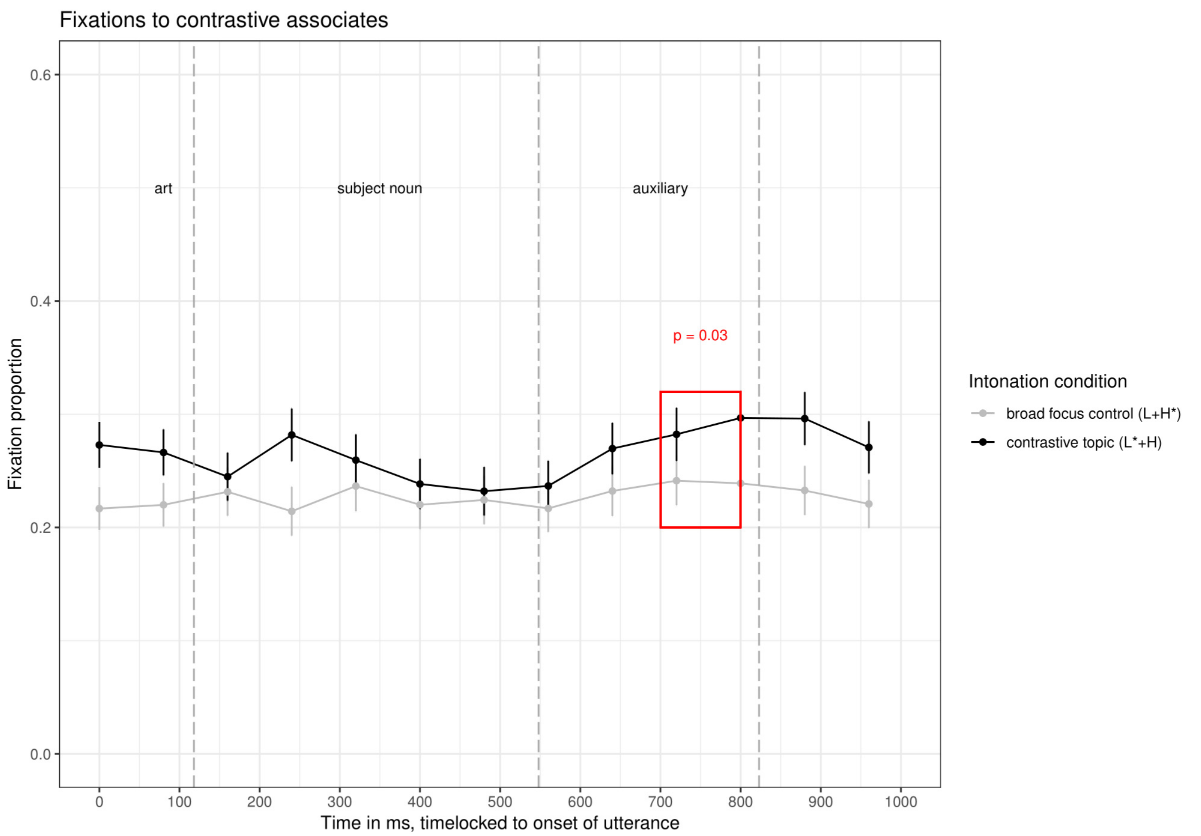


FIGURE 8 | Comparison of fixations to the contrastive associate in the two intonation conditions of Experiment 2.

TABLE 3 | Results of statistical analysis of fixations in subsequent 100 ms time windows for Experiment 2.

	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms
Contrastive associate (Exp2)	**$p = 0.05$**	$p = 0.5$	$p = 0.7$	$p = 0.9$	$p = 0.2$	$p = 0.1$	**$p = 0.03$**
Contrastive associate (Exp1)	$p = 0.1$	$p = 0.1$	$p = 0.07$	$p = 0.1$	**$p < 0.005$**	$p < 0.07$	$p = 0.3$

For comparison, the results of Experiment 1 are repeated in the second row. Bold face indicates a significant difference at alpha < 0.05.

Experiment 1, prenuclear L*+H led to more fixations to the contrastive associate than prenuclear L+H*, even though both contours were manipulated to have the same average F0-excursion in the rise. In combination with the data from Experiment 1 we can conclude that the exact peak height and F0-excursion had no influence on the presence of the effect. Statistically, the effect of intonation contour did not differ across experiments, but it is fair to acknowledge that the fixation differences reached significance later in Experiment 2 than in Experiment 1 (700–800 ms after the onset of the utterance in Experiment 2 compared to 500–600 ms in Experiment 1). Note that the time window at which the effect of intonation contour surfaced in Experiment 2 is one in which the processing of the noun is still taking place. Since the time it takes to plan a saccade is quite variable across listeners (Matin et al., 1993), it is also possible that some participants were already processing segmental information of the auxiliary and intonational information from the F0-transition (high vs. declining). Psychophonetically a high plateau following a rise has been shown to lead to the perception of peak delay (D'Imperio et al., 2010), which is a cue to contrastive topic interpretation, at least in offline studies (Braun, 2004).

We see two possible interpretations for why the effect of intonation occurs a bit later in Experiment 2. First, the resynthesized stimuli in Experiment 2 may take longer to process compared to the natural stimuli in Experiment 1. Previous research has already shown than a resynthesized and unfamiliar intonation contour slows down lexical access (Braun et al., 2011) and this may affect the activation of alternatives as well. This explanation predicts that any kind of unnaturalness in the stimuli leads to later effects, a prediction that can be tested in future experiments. Second, it is possible that – in the absence of a distinctive difference in the F0-excursion of the rise – the pitch accent contrast was blurred in the resynthesized stimuli and that listeners therefore used information on the F0-contour following the stressed syllable (high plateau in the case of prenuclear L*+H and a declining pitch in the case of prenuclear L+H*), a cue that by itself is not distinctive (Braun and Asano, 2013). The F0-information following the accented syllable disambiguates whether the L+H* accent is prenuclear or nuclear. In any case, the fixation data show that listeners activate contrastive alternatives for words produced with a prenuclear L*+H accent even though its acoustic salience was reduced by reducing its F0-excursion (e.g., Mixdorff and Widera, 2001).

We now briefly turn to fixations to the non-contrastive associate. Once again, they did not differ in the two intonation contours, which lends further support to the assumption that only contrastive associates are affected by contrastive pitch accents. Experiment 2 has shown that prenuclear L*+H is among the pitch accents that are processed contrastively, even when this accent had a reduced F0-excursion in the rise.

GENERAL DISCUSSION AND CONCLUSION

Regarding hypothesis H1, which addressed the issue of whether *prenuclear* accents can in principle activate contrastive alternatives, the current fixation data showed that the prenuclear L*+H accents in German do not differ from nuclear focus accents in this respect. Similar to nuclear focus accents, pitch accent type matters for whether or not contrastive alternatives are evoked. In both Experiments 1 and 2, listeners fixated more on contrastive alternatives to the subject noun (e.g., *diver* upon hearing *swimmer*) when it was produced with a prenuclear L*+H accent (which may signal a contrastive topic interpretation) compared to a prenuclear L+H* accent (which is most compatible with a broad focus interpretation). Hence, claims in the literature that prenuclear accents are ornamental, mainly used for rhythmic purposes and remembered and processed poorly (Büring, 2007; Calhoun, 2010; Kapatsinski et al., 2017; Roettger and Cole, 2018) do not hold for all prenuclear accents alike. Clearly, in German, prenuclear L*+H stands out in that respect. From a semantic/pragmatic perspective, this is not surprising, since theories of contrastive topic assume that CT-constituents (marked with prenuclear L*+H in German) evoke alternatives. However, since many of those theories are on English, where the prosodic marking for a CT-interpretation includes a boundary tone (making the accent on the CT-constituent nuclear), it was unclear so far whether this formalization had to do with the fact that contrastive topics are realized with nuclear contours in English, which are known to activate alternatives, or whether it is the result of additional (e.g., syntactic) factors. Our data resolve this issue and indicate that prosodically marked CT-constituents do activate alternatives, even in a language in which CT-constituents are marked by prenuclear accents. In sum, the dichotomy between nuclear and prenuclear does not seem to be very informative for determining which accents are processed as contrastive and which are not.

Regarding H2, which addressed whether or not contrastive topics are processed in the same way as focus constituents with a nuclear L+H* accent on the subject (which were investigated in Braun et al., 2018a), the fixation data clearly show that there is no difference: both the effect size and the timing of the effects were similar. If anything, then the effect is even larger for contrastive topics than for focus constituents, but the cross-experiment comparison was not significant. To corroborate the proposal that contrastive topics behave like focus during online processing, it may be fruitful to investigate other properties that are attributed to the processing of focused constituents. For instance, focused constituents are processed faster than non-focused constituents (e.g., Cutler, 1976; Cutler and Foss, 1977; Cutler and Clifton, 1984), and remembered better (e.g., Gernsbacher and Jescheniak, 1995; Fraundorf et al., 2010). Similarly, while our data does not show a difference in how focus constituents with a nuclear L+H* accent on the subject and CT-constituents evoke alternatives, it would be very important to understand whether speakers treat CT-constituents differently from (standard) narrow focus constituents later on and, in this vein, whether CT-constituents differ from other constituents identified in the literature as more common (aboutness) topic constituents (marked syntactically or morphologically as such, e.g., by left dislocation in German or morphological marking in Japanese). Constituents more standardly understood as (aboutness) topic constituents are, for example, claimed to be better remembered than (standard) narrow focus constituents (see, e.g., Repp and Drenhaus, 2015).

These experimental data shed some light on theories of information structure and of contrastive topic. As outlined before, we take more fixations in the contrastive topic condition relative to the broad focus control condition as indication that the speaker is considering alternatives to the spelled-out element in generating the utterance's interpretation (i.e., the element generating alternatives is). From this perspective, our fixation data show that the processing of CT-constituents is just like that of focus constituents. Given that in CT-constituents, as with focus constituents, alternatives are activated online as the utterance unfolds over time, and that L*+H prenuclear accents indicate CT-interpretations (Braun, 2004, 2005; Braun and Asano, 2013), the results discard incarnations of the "focus within topic" proposals [see (6a) above, repeated here as (11a)] requiring that CT-interpretations are arrived at after full-syntactic processing and identification of the constituent as syntactic topic: given that constituents with L*+H marking evoke alternatives online and are interpreted as contrastive topics, we can discard analyses in which we need to have a full syntactic analysis to then go back and interpret the L*+H constituent as a contrastive topic.

(11) Context question: Was haben die Kinder gespielt? (What did the kids play?)

a. Answer: [[Die [Jungen]$_{\text{Focus}}$]$_{\text{Topic}}$ [haben [Hockey]$_{\text{FOCUS}}$] gespielt.] (focus within topic)

b. Answer: [[Die Jungen]$_{\text{CT}}$ [haben [Hockey]$_{\text{FOCUS}}$] gespielt.] (Büring, 2003, 2016)

c. Answer [[Die Jungen]$_{\text{Focus+computation}}$ [haben [Hockey]$_{\text{FOCUS}}$] gespielt.] (Constant, 2014)

'[The boys] [played [hockey.]$_{\text{FOCUS}}$]'

The results allow for incarnations of the focus within topic theory in which L*+H both marks the constituent as focus and also identifies the constituent as a topic of some sort at the same time. This latter option would be equivalent, on this respect, to considering CT as a basic notion of information structure on its own (11b), and would also be compatible with a notion of CT as focus with special instructions regarding how to manipulate the evoked alternatives in the computation (11c). Regarding the contrast between predictions drawn from Büring's proposal (11b) and those drawn from Constant's proposal (11c), given that the experimental results show that the effect observed in processing CTs is similar to that in processing focus constituents, there is no empirical support from this data to maintain a more complex information structural taxonomy in which CT is different from focus. To be clear, the data does not discard (11b), but if F-marked elements are elements that evoke alternatives relevant for the interpretation and there is no difference between the activation of alternatives for the prenuclear L*+H accent and the nuclear L+H* accent[10], we do not find in these results support for a theory that considers two different notions of information structure, contrastive topic and focus, and a theory that subsumes the two under the same category is more appealing, i.e., (11c).

Our data also speak directly to hypothesis H3, which addressed the role of pitch accent type versus F0-excursion (which is related to intonational prominence) for the activation of alternatives. The fixations in Experiment 2, in which the stimuli were resynthesized so that the prenuclear pitch accents L*+H and L+H* had the same F0-minimum, F0-maximum and F0-excursion, did not differ statistically from those of Experiment 1. This suggests that listeners did not directly react to the F0-excursion of the accents tested but processed the accent type (L*+H vs. L+H*). A closer inspection of the data shows that the effect occurred later in Experiment 2 with resynthesized stimuli than with the natural stimuli in Experiment 1. In section "Discussion" we discussed several options for the later occurrence of the effect in Experiment 2, such as general processing delays with resynthesized or unnatural stimuli as compared to natural stimuli, which are well documented in the literature (Braun et al., 2011). Due to the slightly different timing of the effect, participants had access to the information from the post-stressed syllable, which they lacked in Experiment 1. This may signal the listener whether the prenuclear L+H* accent in the control condition is in fact prenuclear or nuclear, a difference that mattered in Braun et al. (2018a). Although the transition/interpolation of F0 between the prenuclear and nuclear accent did not matter for participants when judging the appropriateness of the intonation contour in different contexts in an offline task, participants may be more affected in an online paradigm. Nevertheless, the available data pose a hen-and-egg problem: We do not know whether our F0-manipulation led to slower processing and hence to the availability of that information or whether the F0-manipulation jeopardized an important aspect of the pitch accent contrast (the onglide, cf. Ritter and Grice, 2015) so that participants had to use information on the F0-movement following the accented syllable. Overall, the data from Experiment 2 are compatible with an interpretation that pitch accent type (signaled by differences in tonal alignment) mattered more for the activation of alternatives

[10]Note that the lack of a difference between the activation of alternatives for the prenuclear L*+H accent and the nuclear L+H* accent is not a null effect in the classical sense of the term. The reason that there is no difference is that both prenuclear L*+H and nuclear L+H* show equally strong fixation differences compared to the broad focus control condition, at the same time window.

than the peak height or F0-excursion of the pitch accents (and the prominence that goes along with these factors, cf. Mixdorff and Widera, 2001). Future studies are necessary to determine the relative strengths of individual prosodic cues that can signal a constituent as CT, also including non-tonal cues such as duration and intensity.

Taken together our findings show that prenuclear L*+H on the sentence-initial subject, an accent that triggers a contrastive topic interpretation of the subject, leads to the activation of alternatives. This is the first study to show that a kind of prenuclear accent immediately evokes alternatives and that differences in accent type (alignment differences) matter for online processing irrespective of peak height or F0-scaling. Generally speaking, it is interesting to note that prenuclear L*+H in Experiment 1 here, but not nuclear H+L* (Experiment 1b in Braun et al., 2018a), activated alternatives to the accented word, since both accent types share a common feature, that the stressed syllable is low-pitched. One explanation is that rising accents of this type are more prominent than falling accents (Baumann and Röhr, 2015; Baumann and Winter, 2018). This asymmetric pattern is mirrored by psychoacoustic studies on just noticeable differences (JNDs) for rising and falling contours (Jongman et al., 2017), who found that English listeners had lower JNDs for rising than falling contours, suggesting a heightened sensitivity for rising contours. This may also hold for German listeners. Yet, the nuclear H+L*, which does not activate alternatives on its own, is the accent that is preferred after a contrastive topic (Braun and Asano, 2013). It is conceivable that the processing of contrastive topic constituents also affects the processing of the subsequent H+L*-marked focus, such that listeners activate alternative to this focus constituent, too. It is an open issue why German, unlike English, identifies contrastive topic constituents with prenuclear and not with nuclear accents and why it uses an accent type (L*+H) that, in nuclear position at least, is judged as less prominent than the prenuclear accent used in broad focus conditions (L+H*). More work is necessary to unravel the effects of pitch and other suprasegmental cues to prominence (Baumann and Winter, 2018) and the role of prominence on the activation of alternatives.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ERB of the University of Konstanz (30/2016). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BB designed Experiment 1, statistically analyzed both experiments, and focussed on processing and prosody. MB and BB designed Experiment 2 together and worked on the Introduction and Discussion sections. MB focussed on semantics and pragmatics.

FUNDING

This work was supported by a grant from the German Research Foundation (DFG) awarded to BB [BR-3428/1-1].

ACKNOWLEDGMENTS

We thank Yuki Asano, Angela James, Verena Köppel and Katharina Zahner for support in stimulus preparation and experiment setup and to Nathalie Czeke, Sophie Egger, Justin Hofenbitzer, Clara Huttenlauch, Jasmin Rimpler, Johanna Schnell, Elena Schweizer and Svenja Willenborg for testing. We are grateful to Johanna Schnell for proofreading.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01993/full#supplementary-material

DATA SHEETS S1–S3 | Sound files for Experiments 1 and 2.

REFERENCES

Altmann, G., and Kamide, J. (2004). "Now you see it, now you don't: mediating the mapping between language and the visual world," in *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, eds J. M. Henderson, and F. Ferreira, (New York, NY: Psychology Press), 347–386.

Arvaniti, A., Ladd, D. R., and Mennen, I. (1998). Stability of tonal alignment: the case of greek prenuclear accents. *J. Phon.* 26, 3–25. doi: 10.1177/0023830915614602

Atterer, M., and Ladd, D. R. (2004). On the phonetics and phonology of segmental anchoring of F0: evidence from German. *J. Phon.* 32, 177–197. doi: 10.1016/s0095-4470(03)00039-1

Baayen, H. R. (2008). *Analyzing Linguistic data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Gann, T. M., and Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychol.* 37, 201–207. doi: 10.1016/j.actpsy.2010.09.011

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models using Eigen and S4. R Package Version 1.0-6.* Availlable at: http://CRAN.R-project.org/package=lme4 (accessed August 21, 2019).

Baumann, S. (2006). *The Intonation of Givenness - Evidence from German*, Vol. 508. Tübingen: Niemeyer.

Baumann, S., and Grice, M. (2006). The intonation of accessibility. *J. Pragmat.* 38, 1636–1657. doi: 10.1016/j.pragma.2005.03.017

Baumann, S., Grice, M., and Benzmüller, R. (2001). "GToBI – a phonological system for the transcription of German intonation," in *Prosody 2000: Speech Recognition and Synthesis*, eds S. Puppel, and G. Demenko, (Poznan: Adam Mickiewicz University), 21–28.

Baumann, S., Grice, M., and Steindamm, S. (2006). "Prosodic marking of focus domains - categorical or gradient?," in *Paper Presented at the 3rd International Conference on Speech Prosody*, (Germany).

Baumann, S., and Röhr, C. T. (2015). "The perceptual prominence of pitch accent types in German," in *Paper Presented at the* 18th International Congress of the Phonetic Sciences, (Glasgow).

Baumann, S., and Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *J. Phon.* 70, 20–38. doi: 10.1016/j.wocn.2018.05.004

Birch, S., and Clifton, C. (1995). Focus, accent, and argument structure: effects on language comprehension. *Lang. Speech* 38, 365–391. doi: 10.1177/002383099503800403

Birch, S., and Garnsey, S. (1995). The effect of focus on memory for words in sentences. *J. Mem. Lang.* 34, 232–267. doi: 10.1006/jmla.1995.1011

Boersma, P., and Weenink, D. (1992-2011). *Praat - Doing Phonetics by Computer (Version 5.0.*33). Amsterdam: Institute of Phonetics.

Braun, B. (2004). "Answers to the perception of thematic contrast and questions regarding the perception of thematic "non-contrast"," in *Paper Presented at the 2nd International Conference on Speech Prosody*, (Nara).

Braun, B. (2005). *Production and Perception of Thematic Contrast in German*. Frankfurt: Peter Lang.

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Lang. Speech* 49, 451–493. doi: 10.1177/00238309060490040201

Braun, B. (2007). "Effects of dialect and context on the realisation of German prenuclear accents," in *Paper presented at the XVIth International Congress of the Phonetic Sciences*, (Saarbrücken).

Braun, B., and Asano, Y. (2013). "Double contrast is signalled by prenuclear and nuclear accent types alone, not by f0-plateaux," in *Paper Presented at the 14th Annual Conference of the International Speech Communication Association*, (France).

Braun, B., Asano, Y., and Dehé, N. (2018a). When (not) to look for contrastive alternatives: the role of pitch accent type and additive particles. *Lang. Speech* 5, doi: 10.1177/0023830918814279

Braun, B., Dehé, N., Neitsch, J., Wochner, D., and Zahner, K. (2018b). The prosody of rhetorical and information-seeking questions in German. *Lang. Speech* 62, 1–5. doi: 10.1177/0023830918816351

Braun, B., Dainora, A., and Ernestus, M. (2011). An unfamiliar intonation contour slows down on-line speech comprehension. *Lang. Cogn. Process.* 26, 350–375. doi: 10.1080/01690965.2010.492641

Braun, B., and Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Lang. Cogn. Process.* 25, 1024–1043. doi: 10.1080/01690960903036836

Breen, M., Fedorenko, E., Wagner, M., and Gibson, E. (2010). Acoustic correlates of information structure. *Lang. Cogn. Process.* 25, 1044–1098. doi: 10.1080/01690965.2010.504378

Büring, D. (1997). *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. London: Routledge.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguist. Philos.* 26, 511–545.

Büring, D. (2007). "Intonation, Semantics and Information Structure," in *The Oxford Handbook of Linguistic Interfaces*, eds G. Ramchand, and C. Reiss, (Oxford: Oxford University Press), 445–474.

Büring, D. (2009). "Towards a typology of focus realizations," in *Information Structure: Theoretical, Typological, and Experimental Perspectives*, eds M. Zimmermann, and C. Féry, (New York: Oxford University Press), 177–205. doi: 10.1093/acprof:oso/9780199570959.003.0008

Büring, D. (2016). "Contrastive Topic," in *The Oxford Handbook of Information Structure*, eds C. Féry, and S. Ishihara, (Oxford, UK: Oxford University Press).

Calhoun, S. (2009). "What makes a word contrastive? Prosodic, semantic and pragmatic perspectives," in *Where Prosody Meets Pragmatics*, eds D. Barth-Weingarten, N. Dehé, and A. Wichmann, (Bingley: Emerald Group), 53–77.

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: a probabilistic perspective. *Language* 86, 1–42. doi: 10.1353/lan.0.0197

Chen, A., Den Os, E., and De Ruiter, J. P. (2007). Pitch accent type matters for online processing of information status: evidence from natural and synthetic speech. *Linguist. Rev.* 23, 317–344.

Chodroff, E., and Cole, J. (2018). "Information structure, affect and prenuclear prominence in American English," in *Paper Presented at the Interspeech* 2018, (Hyderabad).

Cohen, A., and 't Hart, J. (1967). On the anatomy of intonation. *Lingua* 19, 177–192. doi: 10.1016/0024-3841(69)90118-1

Cole, J., Mo, Y., and Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Lang. Cogn. Process.* 25, 1141–1177. doi: 10.1080/01690960903525507

Constant, N. (2014). *Contrastive Topic: Meanings and Realizations*. Ph.D. Dissertation, UMass Amherst, Amherst, MA.

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognit. Psychol.* 6, 84–107. doi: 10.1016/0010-0285(74)90005-x

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Percept. Psychophys.* 20, 55–60. doi: 10.3758/bf03198706

Cutler, A., and Clifton, C. E. (1984). "The use of prosodic information in word recognition," in *Attention and performance X*, eds H. Bouma, and D. G. Bouwhuis, (Hillsdale, NJ: Erlbaum), 183–196.

Cutler, A., and Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Lang. Speech* 20, 1–10. doi: 10.1177/002383097702000101

Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cognit. Psychol.* 42, 317–367. doi: 10.1006/cogp.2001.0750

Dahan, D., Tanenhaus, M. K., and Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *J. Mem. Lang.* 47, 292–314. doi: 10.1007/s10936-014-9306-6

Dennison, H. Y., and Schafter, A. (2010). "Online construction of implicature through contrastive prosody," in *Paper Presented at the Fifth International Conference on Speech*, (Prosody).

D'Imperio, M., Gili Fivela, B., and Niebuhr, O. (2010). "Alignment perception of high intonational plateaux in Italian and German," in *Paper Presented at the 5th International Conference on Speech Prosody*, (Chicago).

Eady, S. J., and Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *J. Acoustic Soc. Am.* 80, 402–415. doi: 10.1121/1.394091

Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., and Lotts, D. W. (1986). Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments. *Lang. Speech* 29, 233–251. doi: 10.1177/002383098602900304

Eberhard, K., Spivey-Knowlton, M., Sedivy, J., and Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *J. Psycholinguist. Res.* 24, 409–436. doi: 10.1007/bf02143160

Esteve-Gibert, N., Portes, C., Schafter, A., Hemforth, B., and D'Imperio, M. (2016). "Intonation in the processing of contrast meaning in French: An eye-tracking study," in *Paper Presented at the International Conference on Speech Prosody*, (Boston).

Féry, C. (1993). *German Intonational Patterns*. Tübingen: Niemeyer.

Firbas, J. (1975). "On the thematic and non-thematic section of the sentence," in *Style and Text: Studies Presented to Nils Erik Enkvist*, ed. H. Ringbom, (Stockholm: Skriptor), 317–334.

Fischer, B. (1992). "Saccadic reaction time: implications for reading, dyslexia, and visual cognition," in *Eye movements and visual cognition*, ed. K. Rayner, (New York NY: Springer), 31–45. doi: 10.1007/978-1-4612-2852-3_3

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *J. Mem. Lang.* 63, 367–386. doi: 10.1016/j.jml.2010.06.004

Gernsbacher, M. A., and Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. *Cognit. Psychol.* 29, 24–58. doi: 10.1006/cogp.1995.1011

Gotzner, N. (2014). *Establishing alternative sets*. Berlin: Humboldt-Universität zu Berlin

Gotzner, N., Spalek, K., and Wartenburger, I. (2013). "How pitch accents and focus particles affect the recognition of contextual alternatives," in *Paper Presented at* *the Proceedings of the 35th Annual Conference of the Cognitive Science Society*, (Austin, TX.).

Grice, M., Baumann, S., and Benzmüller, R. (2005). "German Intonation in Autosegmental-Metrical Phonology," in *Prosodic Typology. The Phonology of Intonation and Phrasing*, ed. J. Sun-Ah, (Oxford: Oxford University Press), 55–83. doi: 10.1093/acprof:oso/9780199249633.003.0003

Gundel, J. (1988). "Universals of topic-comment structure," in *Studies in Syntactic Typology*, Vol. 209-239, eds M. Hammond, E. Moravesik, and J. Wirth, (Amsterdam: John Benjamins).

Gussenhoven, C. (2002). "Intonation and interpretation: Phonetics and phonology," in *Paper Presented at the 1st International Conference on Speech Prosody*, (Aix-en-Provence).

Gussenhoven, C., Repp, B. H., Rietveld, A. C. M., Rump, H. H., and Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *J. Acoustical Soc. Am.* 102, 3009–3022. doi: 10.1121/1.420355

Gyuris, B. (2002). *The Semantics of Contrastive Topics in Hungarian*. Ph.D. thesis, Eötvös Loránd University, Budapest.

Gyuris, B. (2009). Quantificational contrastive topics with verum/falsum focus. *Lingua* 119, 625–649. doi: 10.1016/j.lingua.2007.11.006

Halliday, M. A. K. (1967a). *Intonation and Grammar in British English*. The Hague: Mouton.

Halliday, M. A. K. (1967b). Notes on transitivity and theme in english - part 2. *J. Linguist.* 3, 177–274.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Arnold.

Huettig, F., and McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *J. Mem. Lang.* 57, 460–482. doi: 10.1016/j.jml.2007.02.001

Husband, M. E., and Ferreira, F. (2012). "Generating contrastive alternatives: Activation and suppression mechanisms," in *Paper Presented at the 25th Annual CUNY Conference on Human Sentence Processing*, (New York, NY).

Husband, M. E., and Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Lang. Cogn. Neurosci.* 31, 217–235. doi: 10.1097/WCO.0b013e328316e6e0

Ito, K., and Speer, S. S. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J. Mem. Lang.* 58, 541–573. doi: 10.1016/j.jml.2007.06.013

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jacobs, J. (1982). *Syntax und Semantik der Negation im Deutschen*. München: Fink.

Jongman, A., Qin, Z., Zhang, J., and Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *J. Acoustical Soc. Am.* 142, El163–El169. doi: 10.1121/1.4995526

Kapatsinski, V., Olejarczuk, P., and Redford, M. A. (2017). Perceptual learning of intonation contour categories in adults and 9- to 11-year-old children: adults are more narrow-minded. *Cogn. Sci.* 41, 383–415. doi: 10.1111/cogs.12345

Klein, W. (2008). *The topic Situation. In B. e. a. Ahrenholz (Ed.), Empirische Forschung und Theoriebildung. Festschrift für Norbert Dittmar zum 65. Geburtstag*. Berlin: Peter Lang, 287–305.

Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements on reading. *Eu. J. Cogn. Psychol.* 16, 262–284. doi: 10.1080/09541440340000213

Kohler, K. J. (1991). Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 25, 115–185.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguist. Hung.* 55, 243–276. doi: 10.1556/aling.55.2008.3-4.2

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2016). *lmerTest: Tests in Linear Mixed Effects Models. R package version* 2.0-33. Availlable at: https://CRAN.R-project.org/package=lmerTest (accessed August 21, 2019).

Ladd, D. R., and Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *J. Phon.* 25, 313–342. doi: 10.1006/jpho.1997.0046

Lambrecht, K. (1994). *Information Structure and Sentence Form. Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lommel, N., and Michalsky, J. (2017). "Der Gipfel des spotts. die ausrichtung von tonhöhengipfeln als intonatorishes indiz für sarkasmus [peak alignment as intonational cue to sarcasm]," in *Diversitas Linguarum 42*, eds N. Levkovych, and A. Urdze, (Bremen: Universitätsverlag Dr. N. Brockmeyer).

Matin, E., Shao, K. C., and Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccadic overhead. *Percept. Psychophys.* 53, 372–380. doi: 10.3758/bf03206780

McQueen, J. M., and Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Q. J. Exp. Psychol.* 60, 661–671. doi: 10.1080/17470210601183890

Mehlhorn, G. (2001). Produktion und perzeption von hutkonturen im deutschen [production and perception of hat pattern in German]. *Linguist. Arbeitsberichte* 77, 31–57.

Mixdorff, H., and Widera, C. (2001). "Perceived prominence in terms of a linguistically motivated quantitative intonation model," in *Paper Presented at the 7th European Conference on Speech Communication and Technology*, (Aalborg).

Molnár, V. (1998). Topic in focus - On the syntax, phonology, semantics and pragmatics of the so-called' contrastive topic' in Hungarian and German. *Acta Linguist. Hung.* 45, 89–166.

Morey, R. D., and Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs. R package version* 0.9.12-4.2.

O'Connor, J. D., and Arnold, G. F. (1973). *Intonation of Colloquial English*. London: Longman.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation. Unpublished PhD Thesis*. Bloomington: MIT.

Porretta, V., Kyröläinen, A., Van Rij, J., and Järvikivi, J. (2017). *VWPre: Tools for preprocessing visual world data, R package version 1.0.0*.

Reinhart, T. (1981). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27, 53–94.

Repp, S., and Drenhaus, H. (2015). Intonation influences processing and recall of left-dislocation sentences by indicating topic vs. focus status of dislocated referent. *Lang. Cogn. Neurosci.* 30, 324–346. doi: 10.1080/23273798.2014.923104

Ritter, S., and Grice, M. (2015). The role of tonal onglides in German nuclear pitch accents. *Lang. Speech* 58, 114–128. doi: 10.1177/0023830914565688

Roberts, C. (1996). "Information structure: towards an integrated formal theory of pragmatics," in *Paper Presented at the Papers in Semantics (Working Papers in Linguistics* 49), (Columbus: The Ohio State University).

Roettger, T. B., and Cole, J. (2018). "). On the (ir)relevance of prenuclear accents for intonational learning," in *Paper Presented at the 16th Conference on Laboratory Phonology*, (Lisbon).

Rooth, M. (1985). *Association with Focus*. PhD thesis, University of Massachusetts: Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semant.* 1, 75–116. doi: 10.1007/bf02342617

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguist. Inq.* 31, 649–689. doi: 10.1162/002438900554505

Steube, A. (2001). Grammatik und Pragmatik von Hutkonturen (Grammar and pragmatics of hat patterns). *Linguistische Arbeitsberichte* 77, 7–29.

Strawson, P. F. (1964). Identifying reference and truth values. *Theoria* 30, 96–118. doi: 10.1111/j.1755-2567.1964.tb00404.x

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *J. Acoust. Soc. Am.* 89, 1768–1776. doi: 10.1121/1.401019

Tomioka, S. (2010a). "Contrastive topics operate on speech acts," in *Information Structure: Theoretical, Typological, and Experimental Perspectives*, ed. M. Zimmermann and C. Féry, (Oxford: Oxford University Press), 115–138.

Tomioka, S. (2010b). A scope theory of contrastive topics. *Iberia* 2, 113–130.

Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology* 19, 77–120. doi: 10.1017/s095267570200427x

Vallduví, E., and Engdahl, E. (1996). The linguistic realisation of information packaging. *Linguistics* 34, 459–519.

Vallduví, E., and Vilkuna, M. (1998). "On rheme and kontrast," in *The Limits of Syntax*, eds P. Culicover, and L. McNally, (New York: Academic Press), 79–106.

Wagner, M. (2012). Contrastive topics decomposed. *Semant. Pragmat.* 5, 1–54.

Watson, D., Tanenhaus, M. K., and Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. *L+H**. *Cogn. Sci.* 32, 1232–1244. doi: 10.1080/03640210802138755

Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Lang. Speech* 49, 367–392. doi: 10.1177/00238309060490030301

Wunderlich, D. (1991). Intonation and contrast. *J. Semant.* 8, 239–251. doi: 10.1093/jos/8.3.239

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

TABLE A1 | Subject noun (in German, broad phonetic transcription in IPA and English translation), together with contrastive and non-contrastive associate.

Subject noun	Contrastive associate	Non-contrastive associate
Schwimmer ['ʃvɪṁɐ] (swimmer)	Taucher (53) (diver)	Bad (15) (baths)
Turner ['tʰu͡ɐ.nɐ] (gymnast)	Tänzer (5) (dancer)	Sport (15) (sports)
Nonne ['nɔṅə] (nun)	Mönch (32) (monk)	Kloster (50) (abbey)
Artistin [ʔa͡ɐ.'tʰɪs.tʰɪn] (artist)	Clown (37) (clown)	Zirkus (55) (circus)
Italiener [ʔɪtʰal.'jeː.nɐ] (Italian)	Spanier (37) (Spaniard)	Spaghetti (10) (spaghetti)
Japaner [ja.'pʰaː.nɐ] (Japanese)	Chinese (37) (Chinese)	Asien (10) (Asia)
Kunde ['kʰʊn.də] (customer)	Verkäufer (16) (shop assistant)	Geschäft (30) (shop)
Segler ['zeː.glɐ] (sailor)	Kapitän (21) (captain)	Boot (20) (boat)
Mieter ['miːtʰɐ] (tenant)	Nachbar (32) (neighbor)	Wohnung (35) (apartment)
Professor [pʁo. 'fɛṡo͡ɐ] (professor)	Student (58) (student)	Universität (30) (university)
Schreiner ['ʃʁa͡ɪ.nɐ] (carpenter)	Tischler (11) (cabinet maker)	Holz (40) (wood)
Direktor [di.'ʁɛkʰ.tʰo͡ɐ] (director)	Sekretär (16) (secretary)	Schule (45) (school)
Züchter ['t͡syç.tʰɐ] (breeder)	Bauer (32) (farmer)	Tiere (50) (animals)
Sänger ['zɛŋɐ] (singer)	Techniker (68) (technician)	Lieder (30) (songs)
Maler ['maːl ɐ] (painter)	Zeichner (21) (draftsman)	Farben (30) (paint)
Schlagzeuger ['ʃlaː k.ˌt͡sɔ͡ɪ.gɐ] (drummer)	Gitarrist (21) (guitarist)	Band (45) (band)
Schafe ['ʃaː.fə] (sheep)	Ziegen (21) (goats)	Herde (25) (flock)
Biene ['b̥ iː.nə] (bee)	Wespe (42) (wasp)	Honig (25) (honey)
Flamingo [fla.'mɪŋ.go] (flamingo)	Pelikan (16) (pelican)	Vogel (15) (bird)
Wale ['vaː.lə] (whales)	Haie (16) (sharks)	Orcas (5) (orcas)
Frauchen ['fʁa͡ʊ.ç ən] (mistress)	Herrchen (42) (master)	Hund (45) (dog)
Tiger ['tʰi ː.gɐ] (tiger)	Löwe (58) (lion)	Streifen (10) (stripes)
Rehe ['ʁeː.jə] (deer)	Hirsche (15) (stags)	Wald (20) (forest)
Geiger ['g̊a͡ɪ.gɐ] (violinist)	Pianist (21) (pianist)	Violine (15) (violin)

The number in brackets refers to the percentage of participants that named this associate in the web experiment (N = 19).

TABLE A2 | Mean values and standard deviations in the two intonation conditions of Experiment 1.

	Prenuclear L+H* in broad focus control condition (Experiment 1)	Prenuclear L*+H in contrastive topic condition (Experiment 1)
L-alignment with respect to start of stressed syllable in ms	−25.9 (48.5)	178.0 (57.9)
H-alignment with respect to end of stressed syllable in ms	−45.3 (33.2)	122.5 (50.1)
F0-excursion of the pitch rise in semitones	5.9 (1.1)	8.2 (1.3)
F0-minimum before the pitch rise in Hz	191.4 (8.1)	151.5 (8.5)
F0-maximum after the pitch rise in Hz	287.3 (13.8)	245.7 (8.6)
Duration of the stressed syllable in ms	247.5 (37.3)	272.6 (41.8)
Duration of the subject-NP in ms	421.0 (72.4)	457.8 (77.3)

TABLE A3 | Summary of *p*-values of fixations to the target (Experiment 1).

	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms
Target (Exp 1)	$p = 0.4$	$p = 0.05$	$p = 0.07$	$p = 0.05$	$p = 0.2$	$p = 0.6$	$p = 0.3$

TABLE A4 | Mean values and standard deviations in the two intonation conditions of Experiment 2.

	Broad focus control condition (Experiment 2)	Contrastive topic condition (Experiment 2)
F0-excursion of the pitch rise in semitones	5.7 (2.1)	5.8 (0.6)
F0-minimum before the pitch rise in Hz	184.4 (13.8)	179.3 (6.8)
F0-maximum after the pitch rise in Hz	254.6 (13.8)	249.9 (9.6)

TABLE A5 | Summary of *p*-values of fixations to the non-contrastive associate (Experiment 2).

	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms
Non-contrastive associate (Exp2)	$p = 0.2$	$p = 0.6$	$p = 0.8$	$p = 0.5$	$p = 0.7$	$p = 0.5$	$p = 0.4$

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 October 2019
doi: 10.3389/fpsyg.2019.02184

Investigating the Comprehension of Negated Sentences Employing World Knowledge: An Event-Related Potential Study

Viviana Haase[1]*, *Maria Spychalska*[1,2] *and Markus Werning*[1]

[1] *Institute of Philosophy II, Ruhr University Bochum, Bochum, Germany,* [2] *Department of German Language and Literature I, University of Cologne, Cologne, Germany*

Previous event-related potential (ERP) studies comparing affirmative and negative sentences revealed an N400 for semantically mismatching final words, resulting in a larger N400 for false relative to true affirmative sentences and an opposite effect for negative sentences. Hence, the N400 was independent of the presence of a negation. However, the true negative as well as the false affirmative condition often contained entities or features from different semantic categories and thereby with weak feature overlap, such as e.g., *A cat is (not) a saw* or *Fears are (not) round*, which were then compared to true affirmative and false negative sentences containing entities with stronger feature overlap and partially even hyponomy relations, e.g., *A cat is (not) an animal* or *Planets are (not) round*. Employing world-knowledge variations, in the current study, we investigate whether increasing the feature overlap between the entities of all conditions leads to similar ERP-patterns as in the previous studies. For this purpose, we use sentences of the following type: *George Clooney is (not) an actor* vs. *George Clooney is (not) a singer* where both target words describe a similar profession and thereby function as alternatives to each other. However, in line with the previous studies, we find a truth by polarity interaction, namely, the N400 ERPs are significantly larger for false compared to true affirmative sentences, whereas the effect for negative sentences shows a reversed, though not significant, trend. Overall, the ERP-data suggest that the integration of a negation with the information in its scope is neither fully incremental nor fully delayed, which might be linked to the use of cohyponyms and to the increased feature overlap between alternatives (e.g., *actor, singer*). Additionally, questionnaire-based rating data show that affirmative sentences are perceived as more natural than negative sentences, and, moreover, that true sentences are perceived as more natural than false sentences, independent of their polarity.

Keywords: negation, event-related potentials, N400, world knowledge, language comprehension, alternatives

OPEN ACCESS

Edited by:
Katharina Spalek,
Humboldt University of Berlin, Germany

Reviewed by:
Isabel Orenes,
National University of Distance Education (UNED), Spain
Stephanie Huette,
University of Memphis, United States

***Correspondence:**
Viviana Haase
viviana.haase@rub.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *15 March 2019*
Accepted: *10 September 2019*
Published: *17 October 2019*

Citation:
Haase V, Spychalska M and Werning M (2019) Investigating the Comprehension of Negated Sentences Employing World Knowledge: An Event-Related Potential Study. Front. Psychol. 10:2184. doi: 10.3389/fpsyg.2019.02184

1. INTRODUCTION

Negation is a feature of every human language and an essential element of everyday communication. The addition of a negation operator in a sentence results in a substantial modification of the sentence meaning through a reversal of its truth-value. Despite its frequent use in natural language, the presence of a negative marker seems to elicit additional processing resources during sentence comprehension, resulting in increased reading and reaction times, decreased response accuracy and differential event-related potential (ERP) responses when compared to affirmative sentences (Clark and Chase, 1972; Carpenter and Just, 1975; Fischler et al., 1983; Hasson and Glucksberg, 2006; Kaup et al., 2007; Luedtke et al., 2008; Dale and Duran, 2011; Wiswede et al., 2013; Dudschig et al., 2019). As a consequence, at least when presented in isolation, negative sentences have been argued to constitute an exception to fully incremental language comprehension (Carpenter and Just, 1975; Fischler et al., 1983; Kaup et al., 2006). Incremental comprehension refers to the real-time use of the information in the linguistic input as well as to anticipatory mechanisms regarding upcoming input. Hence, under the assumption of incremental comprehension, the negative marker would have to be integrated in real-time, without delays. While there are circumstances under which negative sentences seem to be processed incrementally—that is if the negation is pragmatically licensed—(Nieuwland and Kuperberg, 2008; Tian et al., 2010; Tian and Breheny, 2015), context-free occurrences of negation are still an open issue with regard to incrementality, and therefore, the general comprehension process of sentences containing a negation operator is still not well-understood. Employing world-knowledge variations, in the current study, we investigate how the use of cohyponyms of a joint hyperonym and thereby a higher overlap of semantic features between the negated entity and its correct alternatives modifies the processing as typically indicated by the N400 ERP, potentially leading to an incremental comprehension process. Additionally, we discuss the role of alternatives in negated sentences. Using context-free sentences furthermore allows a direct comparison to earlier studies employing similar designs.

The N400 component is a negative deflection in the event-related potential that is typically centro-parietally distributed, with a peak around 400 ms after the onset of a stimulus. It is elicited by every content word of a sentence (Kutas and Federmeier, 2000), and its amplitude has repeatedly been shown to be inversely correlated with a word's cloze probability (Kutas and Hillyard, 1984; Gunter et al., 1997; Dambacher et al., 2006; Wlotko and Federmeier, 2012), that is, with the proportion of individuals completing a specific context with that particular word (Federmeier and Kutas, 1999). It was discovered by Kutas and Hillyard (1980) in response to violations of meaning-related expectancy (e.g., *He spread the warm bread with socks*) and has since then been reported in numerous studies (Lau et al., 2009; Kutas and Federmeier, 2011). The size of the N400 has been reported to be modulated by a range of factors such as word frequency (van Petten and Kutas, 1990), atypical thematic role assignments (Weckerly and Kutas, 1999) and plausibility given world knowledge (Van Berkum et al., 1999; Kutas and Federmeier, 2000; Nieuwland and Van Berkum, 2006). More generally, the amplitude of the N400 has been shown to positively correlate with surprisal, that is, the negative logarithm of the conditional probability of the target word given the preceding context (Frank et al., 2015; Kuperberg and Jaeger, 2016). Moreover, Cosentino et al. (2017) and Werning et al. (2019) have shown that it is not only the semantic similarity between the target word and the preceding context (frequency and thematic role assignments held constant) what determines suprisal, but also the relevance of the preceding context for the target word. Furthermore, false compared to true affirmative sentences have repeatedly been shown to lead to an elevated N400 component (Fischler et al., 1983; Hagoort et al., 2004; Nieuwland and Kuperberg, 2008; Metzner et al., 2015; Dudschig et al., 2016, 2019; Spychalska et al., 2016, 2019). For negated sentences, a reversed ERP-pattern has been observed with true negative sentences, such as for example *A rose is not an insect*, eliciting larger N400 components than false negative sentences, such as for example *A rose is not a flower* (Fischler et al., 1983). This interaction of truth-value and sentence polarity also finds support in various behavioral studies (e.g., Clark and Chase, 1972; Hasson and Glucksberg, 2006; Dale and Duran, 2011) as well as in further ERP-studies (Luedtke et al., 2008; Wiswede et al., 2013; Dudschig et al., 2019). As a consequence, it has been assumed that the integration of the negative marker with the information in its scope is not executed in an incremental manner, but instead is a time-consuming process leading to negated sentences requiring additional time to be processed (Carpenter and Just, 1975; Kaup et al., 2006, 2007; Luedtke et al., 2008). It has been argued that comprehending negated sentences requires the initial representation of the underlying affirmative alternative, followed by its integration with the negation and an adapted representation (Carpenter and Just, 1975; Kaup et al., 2006), leading to additional processing time due to the two steps required. For example, to achieve a full understanding of *Barack Obama was not the president of the United States* we need to understand the semantically opposed alternative *Barack Obama was the president of the United States*. In the following, we will provide a brief summary of event-related potential studies employing world knowledge that compare the comprehension of true and false sentences of either affirmative or negative polarity.

In a combined ERP and fMRI study, Hagoort et al. (2004) investigated the integration of different types of knowledge during the comprehension of affirmative sentences (see also Metzner et al., 2015; Dudschig et al., 2016 for recent replications). While false[1] (e.g., *Dutch trains are white and very crowded*) compared to true (e.g., *Dutch trains are yellow and very crowded*) sentences resulted in an N400 effect, semantically incongruent (e.g., *Dutch trains are sour and very crowded*) sentences elicited the highest N400. Furthermore, both false and incongruent sentences led to increased activation of the left inferior frontal cortex. Based on Hagoort et al. (2004)'s findings, the detection of a sentence's falsity and of its semantic anomaly required the same amount of time and activated the same resources. Note, however, that they led to different frequency band activations.

A larger N400 for false (e.g., *A bee is a truck*) compared to true affirmative sentences (e.g., *A bee is an insect*) was also reported by Fischler et al. (1983), who tested affirmative and negative sentences in a truth-value judgment task. Yet, additionally,

[1] *Falsity* means a world knowledge violation, both here as well as in the study by Wiswede et al. (2013) presented below. Instead, *congruence* refers to semantic features, e.g., *sour* is not a feature of trains.

Fischler et al. (1983) reported a larger N400 for true negative (e.g., *A bee is not a truck*) compared to false negative (e.g., *A bee is not an insect*) sentences, hence, the N400 amplitude was independent of the presence of the negative marker as the effect was higher for those sentences where the second noun was semantically unrelated to the first noun and therefore had low feature overlap. Based on these findings, it is difficult to disentangle the effect of truth on the N400 compared to the effect of mere semantic incongruence.

Wiswede et al. (2013) tested whether sentence-related factual world knowledge, e.g., *Yellow is not a number*[2] or *Stones are not soft* is automatically activated as part of the comprehension process and whether it is used to evaluate the truth of affirmative and negative sentences. The participants were split into two groups. Each group had to complete two tasks. Participants of both groups had to respond to a probe task, which consisted of the words "true" and "false" appearing on the screen after 50% of the trials. Participants were asked to press one of two preassigned buttons for each of the two words, respectively. No truth-evaluation was required for this task. The second task varied between groups and occurred after the other 50% of trials. An evaluation group had to respond to a truth-value judgment task, while a control group had to indicate whether a probe sentence was identical to the stimulus sentence or not. In the analysis, ERPs time-locked to the onset of the final word of each sentence from both groups were included, independent of the task. *Group* was added as a separate factor. Wiswede et al. (2013) reported an interaction of truth-value and sentence polarity, that is, a larger N400 for conditions containing a semantic mismatch between subject and object of a sentence (i.e., false affirmative and true negative). This effect occurred in both groups but was stronger in the group who had to complete a truth-value judgment task than in the control group. They interpreted this effect as an indication that the analysis of word meaning and of semantic relations between words within a sentence occurs automatically, independent of the task. Furthermore, they reported significantly stronger N400 amplitudes for negative compared to affirmative sentences for both groups, independently of the sentence truth-value. Additionally, they observed a late negativity for false compared to true sentences in the truth evaluation group. This negativity occurred in a time window between 500 and 800 ms for affirmative sentences, but only later, between 800 and 1,000 ms, for negative sentences. Wiswede et al. (2013) concluded that truth validation is not fully automatic but goal dependent. However, this conclusion was based on a null result in the control group.

Nieuwland and Kuperberg (2008) addressed the interplay between pragmatic context and negation. Participants were presented with affirmative or negative sentences that were either true or false with respect to world knowledge and that were either embedded in a pragmatic context (pragmatically licensed, e.g., *With proper equipment, scuba diving is/isn't very safe/dangerous and often good fun*), or were presented without pragmatic context (pragmatically unlicensed, e.g., *Bulletproof vests are/aren't very safe/dangerous and used worldwide for security*). For the pragmatically unlicensed conditions, the authors observe a larger N400 for false affirmatives, false negatives and true negatives compared to true affirmatives. Hence, they did not observe an effect of truth-value in pragmatically unlicensed sentences on the N400 neither, which matches results from previous studies. For the pragmatically licensed conditions, however, they observed a higher N400 for false affirmative and false negative compared to true affirmative and true negative sentences. These results suggest that negation is implemented into the sentence-level meaning in an incremental manner at least if the negation is pragmatically licensed.

In a very recent study, Dudschig et al. (2019) investigated whether additional time to process the negation operator facilitates its integration into the sentence-level meaning. They compared correct (i.e., true and congruent) sentences to sentences containing either an incongruence or a world-knowledge violation, thus, their design resembled the one by Hagoort et al. (2004). In addition, they tested sentences containing a negation as well. In the first experiment, the negative adverb *nicht* ("not") was placed within the sentence (e.g., *Zebras/Ladybirds/Thoughts are (not) stripy*). In the second experiment, an external negation that takes scope over the whole sentence was tested (e.g., *It is (not) true that zebras/ladybirds/thoughts are stripy*). The idea behind prepending the negation was to give the reader more time to process and integrate it with the information in its scope. The authors reported an N400 for the two violation conditions (incongruence and world knowledge) compared to the correct condition, both for affirmative and negative sentences in both experiments, that is, independent of the position of the negation operator. Therefore, prepending the negation operator to the beginning of the sentence did not facilitate an incremental interpretation[3]. Taken together, the results presented above suggest that the comprehension process of negated sentences is not fully incremental, not even if the system is given additional time to integrate the prepended negation with the information within its scope. As soon as a negation is pragmatically licensed, however, the comprehension process seems to function fully incrementally (Nieuwland and Kuperberg, 2008, see also Tian and Breheny, 2015).

The difficulty in achieving an overall interpretation of earlier studies is the rather strong variation of feature overlap and semantic category mismatch across conditions. For example, Fischler et al. (1983) made use of hyponomy relations such as for example in *A hammer is (not) a tool* which were compared to sentences as for example *A hammer is (not) a fish*, resulting in a mix of semantic categories as well as in a comparison of animated and not animated entity sets. Adding a negation to these sentences results in a true but pragmatically odd sentence compared to a false but in some contexts presumably acceptable

[2]The experiment was done in German, the original sentences were for example *Gelb ist keine Zahl* and *Steine sind nicht weich*. Hence, negation was either marked using the negative quantifier *kein* ("no") or the negative adverb *nicht* ("not") which have different syntactic structures and different scope.

[3]However, providing additional time after the sentences (in a sentence-picture verification paradigm with the sentence preceding the picture) seemed to allow a successful integration of the negation (Kaup et al., 2006, 2007; Luedtke et al., 2008, see also Ferguson et al., 2008).

sentence. The stimuli of Wiswede et al. (2013) show similar problems of semantic category mismatch and animacy violations. For example, they compared sentences like e.g., *Socrates is (not) a country* or *Iron can (not) fly* to sentences like *Five is (not) a number* or *Elephants are (not) small*. Additionally, those sentences that had a noun phrase as target word were preceded by the indefinite article which was, due to the word-by-word presentation, presented in isolation. De Long et al. (2005) reported that readers were able to predict specific words based on the prior occurrence of the indefinite article and the distinction between *a* or *an*, which was either followed by a word beginning with a vowel or a consonant in English (De Long et al., 2005, however, see Ito et al., 2016, 2017 for a debate regarding the replicability of these results). In German, due to grammatical gender, a similar differentiation is possible between *(k)ein* (neutral), *(k)einen* (male) and *(k)eine* (female). Accordingly, the use of stimuli in Wiswede et al. (2013) might have narrowed down the number of potential alternatives, thereby facilitating the anticipation of upcoming words in some trials, leading to heterogeneous material. In the current study that focuses on the comprehension of negated compared to affirmative sentences using world-knowledge, we avoid mixing semantic categories as well as animacy violations. Instead, the current study uses a true description of a publicly well-known person for the true affirmative condition, e.g., *George Clooney is an actor* which is then compared to a false version, e.g., *George Clooney is a singer*[4]. Importantly, the false version was created by using a different profession of public life denoted in a noun as well, thereby increasing the overlap of semantic features compared to the respective true sentence. Additionally we aimed at increasing this overlap by avoiding combinations of professions from rather unrelated fields, e.g., religion and sports. For the negative sentences, the adverbial negative marker *not* is added to these sentences, resulting in a false, e.g., *George Clooney is not an actor*, and a true sentence, e.g., *George Clooney is not a singer*. Using cohyponyms of the hyperonym "profession" across all conditions and increasing the feature overlap between the critical words across conditions, e.g., *actor, singer*, we aim at maximizing the coherence of all sentences to investigate how it affects the comprehension process of negated sentences.

In everyday conversation, negation does not only create a semantic opposition, but furthermore, it licenses the truth of alternatives. In isolated negated sentences, anticipating upcoming content is relatively difficult since the set of true sentence continuations for a negative sentence is vast compared to the relatively small set of true sentence continuations for an affirmative sentence. Logically, every member of the set of *not p* is a potential alternative to *p*. However, during the fast and efficient process of language comprehension, anticipating all potential alternatives would be costly for the cognitive system and furthermore would be highly inefficient since it requires maintaining an infinite amount of alternatives. In principle, potential alternatives can be found along various dimensions, depending on the type of verb that is used and depending on the scope of the negation. For example, in a sentence like *Rachel did not bake the bread*, potential alternatives for the negation can be found along the dimension of the actor, along the dimension of activities and along the dimension of the patient, that is, Rachel could have baked something else, e.g., a cake, she could have done something else to the bread, e.g., cut it, or someone else could have baked the bread[5]. As the example demonstrates, alternatives are semantically related to the negated information (e.g., entity, event). Here, we focus on the dimension of professions that are denoted as nouns in our design. In the above example, reading a sentence fragment like *George Clooney is…*, the reader may anticipate potential content related to this specific person, his profession, career, success, resulting in an expectation of words like e.g., *actor, successful, rich, famous,…*. Instead, for the respective negated sentence *George Clooney is not…*, in theory, every alternative that would make the affirmative sentence false could be anticipated. As a result, the reader might find herself in a situation of not being able to anticipate anything if presented with such a sentence in isolation. However, not all content is equally likely to occur, that is, some potential alternatives are more likely to occur than others. Due to the contextual invariance of negation (Mohammad et al., 2013; Kruszewski et al., 2016), that is, negations typically occurring in the same contexts as their affirmative counterparts, a certain feature overlap between the true continuations for an affirmative sentence and true continuations for its negated counterpart can be assumed. Accordingly, cohyponyms are straightforward alternative candidates. However, some cohyponyms, e.g., professions in the above example, seem more suitable for the negative sentences, than others. Categorization research suggests that many human categories are taxonomic, that is, items are grouped together on the basis of shared perceptual and functional features (Kay, 1971; Rosch et al., 1976). Membership within a category is gradual, determined by whether and how many features an item shares with other members of a category (e.g., Rosch, 1973, 1975). Assuming that we anticipate potential alternatives during online sentence comprehension, a gradual spread of activation in a semantic network can be assumed, in which the level of activation depends on the overlap of features. For example, other related professions as e.g., *a singer, a stage director, a producer* intuitively seem more plausible as a continuation of *George Clooney is not* than less related professions, e.g., *an architect, a pharaoh, an astronomer* would be, and certainly seem more plausible than true "out of category"-alternatives, e.g., *a bread, a dog, a hammer* that are not cohyponyms.

Our sentences are all of the form *X war einmal/nicht Y in Z* (X was once/not Y in Z) or *X ist derzeit/nicht Y in Z* (X is currently/not Y in Z) where *X* denotes a publicly well-known person, *Y* is a noun referring to a profession, and thus, is a

[4]Our experiment is done in German and does not require any article in this construction to yield a grammatical sentence. Thereby, the potentially facilitated predictability just described for the study by Wiswede et al. (2013) does not apply to our design.

[5]This example refers to a sentence presented in isolation and in written language. Context as well as prosody would mark the focus of the sentence, thereby limiting the number of alternatives.

TABLE 1 | Example of the experimental conditions in German with English translation.

	True	False
Affirmative	George Clooney ist derzeit Schauspieler in den USA. George Clooney currently is an actor in the USA.	George Clooney ist derzeit Sänger in den USA. George Clooney currently is a singer in the USA.
Negative	George Clooney ist nicht Sänger in den USA. George Clooney is not a singer in the USA.	George Clooney ist nicht Schauspieler in den USA. George Clooney is not an actor in the USA.

cohyponym of the hyperonym "professions" and is the target word in this experiment, and *Z* refers to a location which can be either a country, a city or a region. The sentences in our experiment all have an SVO-structure with V being the simple past or simple present of the verb *sein* ("to be"). In the negative sentences, the negative adverb *nicht* ("not") was placed between the verb and the object, which is the unmarked position for the negative marker in German. To keep sentence length equal between conditions, the adverbs *einmal* ("once") or *derzeit* ("currently") were inserted into the affirmative sentence, depending on its tense (see Dudschig et al., 2019 for a similar procedure). Hence, the two factors are *Polarity* (affirmative, negative) and *truth-value* (true, false) resulting in a 2 × 2 design. The final prepositional phrase did not alter the truth-value of the sentences and was added to avoid an overlap of effects elicited by the manipulation in the design that could be overlapping with a potential sentence final wrap-up effect. Wrap-up effects in reading are assumed to reflect increased processing associated with intra- and inter-clause integration (Just and Carpenter, 1980; Rayner et al., 2000; Hirotani et al., 2006; Warren et al., 2009). An example of the four conditions is given in **Table 1**. Each sentence was followed by a probe word for which participants had to decide whether it was contained in the previous sentence or not. Employing a probe verification task instead of a truth-value judgment task allows to avoid a potential confound of effects resulting from mere sentence comprehension with effects elicited by the engagement in explicit truth-value judgment. At the same time, the task is more natural than explicit truth-value judgment and requires participants to pay attention to the sentences. Additionally, the type of world knowledge violations we use might sometimes be difficult to be evaluated with a 100% certainty. For example, *George Clooney is not a singer* might seem intuitively correct in terms of world knowledge. However, strictly speaking, to be able to evaluate the truth of this sentence, we would have to have more knowledge about this person to assess whether, e.g., in private, he likes to sing. Such knowledge, however, is not relevant for the current task and is not at the focus of this experiment.

While the overall design of our study resembles the design of the experiments by Dudschig et al. (2019), Fischler et al. (1983), and Wiswede et al. (2013), there are various differences between them. As described above, in contrast to other studies, but especially in contrast to Fischler et al. (1983), across all conditions we use cohyponyms of the hyperonym "profession" thereby avoiding a mix of semantic categories and increasing the overlap of features between entities across conditions. Hence, we use animate entities only. In the experiment by Wiswede et al. (2013) the target word consisted either of an adjective, a noun or a noun phrase and it was preceded by either adverbial negation *nicht* ("not") or quantifier negation *kein* ("no") which have different scope. Instead, our target word is always a noun and the negative marker does not vary. In contrast to Dudschig et al. (2019), in our study, it is the critical word itself that alters between conditions, whereas in their study, it was the first noun of a sentence that differed while the critical word was identical in all conditions.

Semantic knowledge, world knowledge and language comprehension in general are subject to individual differences. For example, a number of studies reported an absence of predictive processes under certain circumstances, e.g., in children with low vocabulary scores (Borovsky et al., 2012), in older persons (Federmeier et al., 2002; Federmeier and Kutas, 2005; DeLong et al., 2012; Wlotko et al., 2012), in second language learners (Martin et al., 2013) and schizophrenic patients (Kuperberg, 2010). While such findings may suggest that certain speaker groups do not engage in predictive processing, it might be possible as well that these speakers anticipate upcoming input during comprehension, but that some of the computations involved are still incomplete when the relevant input arises (Chow et al., 2018). An incomplete computation of the negated sentence meaning that is only completed later in time is consistent with previous studies by Kaup et al. (2006) and Luedtke et al. (2008), as well as Dudschig et al. (2019). Here, we were interested in a potential correlation of working memory capacities and the seemingly time-consuming comprehension process for negated sentences. The high variability of the results of individual subjects reported in Fischler et al. (1983) further motivates controlling for a correlation of individual factors on ERP-results.

Despite the use of cohyponyms and thereby the increased feature overlap of the target words between conditions of our design, we still expect a larger N400 for false compared to true affirmative sentences, in line with earlier experiments (Fischler et al., 1983; Hagoort et al., 2004; Wiswede et al., 2013; Metzner et al., 2015; Dudschig et al., 2016, 2019). This comparison functions as a control comparison, that is, a complete absence of this effect might suggest that the high feature overlap resulted in a similarity between the true and false alternatives that was too strong to be noticed immediately. Furthermore, we hypothesize the modulation of feature overlap between the negated noun and its alternatives to facilitate an incremental comprehension process by increasing the chances to anticipate upcoming content in the negated sentences as well. Here, we use the term *anticipation* to refer to a potential pre-activation of upcoming content in the linguistic input as a result of overlapping features with previously encountered material already processed. The present study neither aims at investigating the automaticity of this process, nor at explicitly tackling the question of whether the N400 reflects expectancy, prediction or integration. As mentioned earlier, in a semantic network that is organized in taxonomies, a gradual spread of activation can be assumed, depending on the overlap of features between words. Due to the increase of overlapping features between the negated noun and potential alternatives we expect a facilitation of the comprehension process. Therefore, we hypothesize a smaller gap between the reaction times and response accuracies as well as reduced N400 effects. Furthermore, the inversion of the N400 for negated sentences, with true negated sentences eliciting higher N400s than false negated sentences (Fischler et al., 1983; Wiswede et al., 2013; Dudschig et al., 2019) might be changed, resulting in a larger N400 for false compared to true negative sentences. If the modulation of feature overlap between true and false alternatives does not affect the processing, we expect our results to match earlier studies (Fischler et al., 1983; Wiswede et al., 2013; Dudschig et al., 2019), hence, then we expect a larger N400 for false compared to true affirmative sentences, and a larger N400 for true compared to false negative sentences. Additionally, we hypothesize a correlation with working memory capacities, resulting in lower N400 effects for people with low working memory capacities. Engaging in anticipatory mechanisms can be expected to be more difficult with comparatively low capacities to store this information, potentially leading to a reduction or even absence of such mechanisms. Instead, high working memory capacities may enable the pre-activation of a range of alternatives, resulting in stronger N400s for each of them.

2. METHOD

2.1. Participants

Thirty-six (fifteen male) students of local universities participated in the experiment (age: 18–38, mean: 26.44, *SD*: 4.31) and were reimbursed for their participation or received course credit. All participants were right-handed monolingual German native speakers who were born in Germany and grew up there. The latter selection criterion was applied to increase the likelihood that they are familiar with the names used in the stimuli sentences. Part of the names are nationally well-known, e.g., due to activities on TV in Germany or in German politics, but not necessarily internationally well-known. They all had normal or corrected-to-normal vision and no history of psychological or neurological problems.

2.2. Material

We created 40 pairs of professions (e.g., *actor*, *singer*). Out of these 40 pairs we created 40 stimuli sets consisting of two true (at the time of data collection) and two false sentences by adding a negation into two of them, hence, the stimuli material consisted of 160 sentences. To avoid repetition effects and direct contradictions within the material, we split the sentences into two lists. Each participant saw only one of the lists. To do so, each quadruple of sentences was assigned two celebrities that matched the true affirmative version equally, thus, a total number of 80 celebrity names was used within the stimuli sentences. In one list, the true affirmative and the true negative of a set were assigned one person (e.g., *George Clooney*), while the false affirmative and the false negative were assigned another person (e.g., *Angelina Jolie*)[6]. With this division, we avoided the contradiction between true affirmative (e.g., *Angelina Jolie is an actor*) and false negative sentences (e.g., *Angelina Jolie is not an actor*) within one list (see e.g., Yurchenko et al., 2013 for a similar procedure). Hence, each critical word appeared four times within a list, once per condition and twice with the same name. Target words had a mean frequency of 11.49 ($SD = 2.55$, range 7–18)[7]. Within one quadruple, the mean difference in frequency was 2.55, $SD = 1.72$. When creating the stimuli, we checked LSA-values (Landauer et al., 1998) of the English translation of the two profession-hyponyms (e.g., *actor/singer*) of each set of conditions[8]. Since both were combined with the same person, the information obtained helped to approximate conceptually related professions and to combine them accordingly. Across conditions, our LSA-values are within the range −0.03 to 0.37 (with one outlier at 0.67), mean = 0.14, $SD = 1.40$.

The material has been created with the help of a questionnaire which was completed prior to the experiment. The online questionnaire consisted of two different parts and was completed by 45 German speakers (21m, age range 19–34 years, mean: 24.46, SD: 4.99) who were born in Germany and grew up there. None of them participated in the EEG experiment. In the first task, participants had to rate how well they know[9] the person whose name was shown to them one by one. They were asked to indicate their response on a four point Likert scale and they were informed in the following way about the scale: 4 = you know a person's name and profession; 3 = you know to whom the name refers to but have little knowledge about that person, e.g., you roughly know that somebody is from politics; 2 = you hardly know a person, that is, you heard the name before but do not know who that person is; 1 = you do not know a person at all. An example was used to demonstrate the distribution of the scale. The questionnaire contained a selection of 113 female and male publicly well-known persons, covering the categories film, sciences, humanities, arts, music and politics in both past

[6]In German, professions usually are inflected for gender, therefore, the noun denoting the profession was adapted accordingly, resulting in e.g., *Angelina Jolie ist Schauspielerin*.

[7]Leipzig Wortschatz http://wortschatz.uni-leipzig.de/de

[8]Obtaining LSA-values for the two nouns within one stimulus sentence in the current design is problematic because of the use of proper names. Not all names can be found in the LSA-databases and they are at risk to lead to a distorted picture. For example, by checking the LSA-value for "George Clooney" and "actor", all entries where he is referred to by "Mr. Clooney" or "Clooney" are ignored. These variations, however, are frequently used to avoid repetitions in texts and therefore are likely to occur in the text pool underlying the database. Furthermore, frequent proper names might be subject to ambiguities within these databases.

[9]They were informed that knowing a person here does not mean knowing them personally but rather knowing who this person is and what this person's field of publicly known profession is.

and present. For the material of the ERP-study, only those names that got a mean of 3 or higher were included in the experimental material. In total, 57 names were selected from the questionnaire [mean across all selected names = 3.63 (range 3.05 − 4), mean $SD = 0.62$ (range 0−0.96)]. Due to the unexpectedly high number of names that needed to be excluded due to mean values lower than 3, additionally 23 names were included that were not rated in the questionnaire[10].

In the second part of the questionnaire the participants saw the same celebrity names in a sentence completion task. They read the beginning of a sentence of the type *X war* (X was) or *X ist* (X is), depending on whether the person is still alive and still active in their profession, where *X* is the name of a person. They were instructed to fill in a noun that they think best describes the publicly known profession of that person. The profession that is mentioned in our true affirmative sentences is the profession that the person is on average mostly known for. We took into account the answers given in the second part of the pre-test questionnaire as well as synonyms, hyponyms and hyperonyms (e.g., *artist*, *painter*) to these answers. When creating false sentences attention was paid to the semantic relatedness. In false affirmative sentences, the profession was taken either from the same or from a close semantic field, e.g., music and arts. We avoided to combine a person with a totally unrelated profession. For example, when creating a false sentence for a musician, a profession from the field of arts (e.g., music, painting, film) was chosen rather than a profession from politics. Furthermore, we avoided combining a person from the past with a relatively modern profession (e.g., show master). Predominantly or exclusively male professions (e.g., dictator, Pope) were not assigned to female names. The negative sentences were derived from the affirmations by adding the negative marker *nicht* ("not").

All sentences ended with a prepositional phrase specifying the true origin of the subject of the sentence (e.g., *from Spain, in Rome*). The verb of the sentence was either the simple present or the simple past of *sein* ("to be"), depending on whether the person is still alive and still active in that field. To keep the sentence length stable and to make affirmative and negative sentences fully comparable, an adverb was inserted in the affirmative sentences after the verb (hence, at the position where the negation is located in the negative sentences). For the sentences using simple present the adverb *derzeit* ("currently") was used, for those sentences using simple past the adverb *einmal* ("once") was used. Those adverbs were chosen as fitting best as a counterpart to the negative marker and were closest in frequency[11] compared with the negative adverb.

Additionally, a total of 76 filler sentences was included to increase the variability of the material. They all had the same structure as the stimuli sentences, but used different professions and other names, including cartoon figures as well. The adverb in the affirmative sentences was varied [*eigentlich* ("actually"), *offenbar* ("obviously"), *bekanntlich* ("as is known"), *damals* ("back then"), *heute* ("today")]. The 76 sentences resulted from 19 quadruples each consisting of true and false affirmative and negative sentences. Hence, the overall distribution of affirmative and negative, true and false sentences was not altered by the fillers.

For the probe task a word in capital letters appeared on the screen and subjects had to decide by pressing a button whether this word was contained in the previous sentence or not. In 50% of the trials the probe word was part of the previous sentence. Words from all sentence positions were pseudo-randomly used in the probe task to avoid that participants would selectively focus on specific words due to the task. In case of incorrect probes, words of the same grammatical categories were used.

The stimuli were rated in two *post-hoc* online-questionnaires regarding their perceived naturalness and their perceived truth-value. The material was split into two lists, as described above for the experiment. The first questionnaire consisted of list A for the naturalness ranking and list B for the truth-ranking, while the second questionnaire consisted of list B for the naturalness ranking and list A for the truth-ranking. Hence, each questionnaire consisted of 320 questions, split into two sections with different rating tasks. The main purpose was again to avoid repetitions and contradictions within one list. The questionnaire was designed in Qualtrics and distributed via the platform Prolific, where participants received payment for their participation. Selection criteria regarding native language, provenience and age were kept identical to those for the ERP-study. In the first part of each questionnaire, participants were asked to rate each sentence regarding its naturalness on a 4-point Likert-scale (4 = natural, 3 = rather natural, 2 = rather unnatural, 1 = unnatural). They were informed that naturalness is not necessarily correlated with truth-value, and that they are therefore allowed to rate false sentences as natural and true sentences as unnatural, if necessary. In the second part, participants were asked to rate each sentence regarding its truth vale, again on a 4-point Likert-scale (4 = true, 3 = rather true, 2 = rather false, 1 = false). We asked them to complete the questionnaire without help. Within one part, the order of sentences was randomized. Each questionnaire was completed by 40 participants, that is, 80 participants (53 male) rated the material in total (mean age 24.68 years, $SD = 4.58$, range 18−39); participants who took part in the first questionnaire were not allowed to complete the second questionnaire.

2.3. Procedure

Upon arrival and after being informed about the procedure of the experiment the participants signed a consent of participation in accordance with the Declaration of Helsinki. Afterwards, participants filled in a translated version of the *Edinburgh Handedness Inventory* test (Oldfield, 1971), and a demographical questionnaire asking for age, handedness, education, vision, medication and neurological and psychological history. Subsequently, they completed two pretests which are part of the WAIS such as a computerized version of the Reading Span (van den Noort et al., 2008) and the Digit Span forward and backward. They furthermore filled in the *Autism Spectrum Quotient Questionnaire* (AQ) (Baron-Cohen et al., 2001), which is

[10] See section 3 for their mean values from a post-experiment rating.

[11] Leipzig Wortschatz http://wortschatz.uni-leipzig.de/de: einmal (6), derzeit (7), nicht (2).

a self-assessment questionnaire that measures traits of the autistic spectrum disorder (ASD) in healthy adults with normal IQ, such as social skills, communication skills, imagination, attention to detail, and attention-switching, which have been reported to be correlated with differences in language comprehension, especially when comparing underinformative to informative sentences (see for example Nieuwland et al., 2010, but see also Spychalska et al., 2016)[12].

The EEG-measurement was conducted in an electrically and acoustically shielded cabin. Participants were seated in front of a screen and a *Cedrus* response box with five buttons out of which the right and the left button were needed for the responses. After the preparation of the electrode cap subjects were given a written instruction and consecutively did a training session consisting of seven example trials. The experiment was programmed in Presentation. No feedback was given throughout the experiment. Participants were asked to attentively read the sentences and respond to the probe task. The experiment was divided into six blocks with breaks in between. The net measurement time was approximately 45 min.

The sentences were displayed on the screen in word-by-word manner in black color against a gray background (to avoid strong contrast, see Gunter et al., 1999). Each trial began with a fixation cross that was presented for 800 ms. The name was presented for 600 ms, followed by a 200 ms blank screen. The verb as well as the negation/adverb were each presented for 400 ms with a 400 ms blank each. The target word as well as the final phrase were each presented for 500 ms. After the target word, the blank lasted for 500 ms, after the final phrase until the occurrence of the probe word the blank lasted 1,000 ms. The probe word was presented maximally 3,000 ms. To respond to the probe verification task participants had to press a button; the probe word disappeared as soon as the participant clicked a response button. The assignment of the right and left button for true and false responses was counterbalanced across subjects. All participants remained naive regarding the purpose of the study.

To assess the participant's knowledge about the stimuli used in the experiment, after they completed the experiment, they filled out a digital questionnaire. It was designed in the same way as the pre-test (see section 2.2) and included every name used in the stimuli sentences (i.e., 80 names).

2.4. EEG Recording and Preprocessing

The EEG was recorded with a 64 channel ActiCap system by BrainVision, band-pass filtered at 0.01-250 Hz and sampled with a frequency of 500 Hz. AFz served as Ground, FCz as physical reference during the recording. To control the vertical and horizontal eye movements four electrodes (FT9, FT10, PO9, and PO10) were removed from their determined location and were placed over and under the right eye as well as on both temples to measure the electrooculogram (EOG). All impedances were kept below 5 kΩ. The data was processed using the Brain Vision Analyzer 2.1 software. We applied an offline band-pass filter of 0.1-30 Hz. All trials with an absolute amplitude difference higher than $200\,\mu V/200\,ms$ or with an activity lower than $0.5\,\mu V$ in intervals of 100 ms or longer were automatically rejected. The maximal allowed voltage step was $50\,\mu V/ms$. Eye-blinks and eye-movements were corrected by a semi-automatic independent component analysis. The data was re-referenced to the linked mastoids (TP9, TP10) and then segmented into epochs of 1,000 ms, beginning at the onset of the second noun, with a −200 ms baseline. The baseline correction serves to remove differences due to drifts, while avoiding a distortion of the post-stimulus ERPs that might result from transient differences between conditions in the baseline interval (Wolff et al., 2008). Before averaging, any segments with remaining physical artifacts lower than $-90\,\mu V$ or higher than $90\,\mu V$ were removed. Across subjects, the minimum of preserved segments was 25 out of 40, however, for most subjects, at least 30 segments per condition (i.e., at least 70%) were preserved. Four participants had to be excluded from the ERP-data analysis due to excessive artifacts leading to a loss of more than 50% of segments per condition for three of them, and due to strong signal drifts on multiple electrodes in the fourth participant. One data set was excluded due to a technical problem during recording, hence the ERP-analysis is performed on 31 participants.

3. RESULTS

3.1. Behavioral Responses to the Probe Task

The mean accuracies and mean response times to the probe task for 31 subjects are shown in **Table 2**. For the behavioral responses the non-parametric Friedman test, which, unlike ANOVA, can be used for samples that are not normally distributed, indicated that the mean accuracy to the probes differed across the four conditions: $\chi^2(3) = 12.457, p = 0.005$ $(N = 31)$. Based on the Wilcoxon *post-hoc* analysis, the effect results from a lower mean accuracy for false negative compared to false affirmative sentences ($z = -2.310$, $p = 0.019$) and from lower mean values for false negative compared to true negative sentences ($z = -2.118, p = 0.033$).

For the reaction times the parametric repeated measures ANOVA revealed an interaction *Polarity*Truth* [$F(1, 30) = 13.758, p = 0.001$, partial $\eta^2 = 0.314$] but no main effect for *Polarity* ($F > 0.05, p > 0.5$) and no main effect of *Truth* ($F > 3, p > 0.08$). We broke down the interaction by *Polarity*. For affirmative sentences, the ANOVA shows a main effect for *Truth* [$F(1, 30) = 13.290$, $p = 0.001$, partial $\eta^2 = 0.307$], with responses to false probes on average taking

TABLE 2 | Mean accuracy in the Probe Verification Task in percentage and mean reaction times in milliseconds for all four conditions; standard deviations are indicated in brackets.

	Accuracy		Reaction time	
	True	False	True	False
Affirmative	97.99 (2.6)	97.61 (3.04)	885.01 (204.567)	924.173 (229.216)
Negative	97.35 (3.00)	96.32 (2.88)	911.737 (204.48)	896.52 (195.77)

N = 31.

[12] We therefore considered it worth testing whether the way in which individuals process negated sentences, which are considered to be underinformative when presented in isolation, correlates with their AQ-score.

longer than responses to true probes ($\Delta_{(False,True)} = 39.16$ ms). There was no effect for negative sentences ($F > 0.2, p > 0.1$).

3.2. ERP-Results

The ERPs elicited by the target word of the sentence were evaluated in a repeated measures ANOVA with *Polarity* (affirmative/negative) and *Truth* (true/false) as within-subject factors. To analyze possible interactions with electrode positions, *Lateralization* (left/right) and *AP* (anterior/posterior) were involved as further factors, resulting in four regions of interest (ROI). Each ROI comprised 11 electrodes: left anterior (FP1, AF3, AF7, F1, F3, F5, F7, FC1, FC3, FC5, FT7), right anterior (FP2, AF4, AF8, F2, F4, F6, F8, FC2, FC4, FC6, FT8), left posterior (CP1, CP3, CP5, TP7, P1, P3, P5, P7, PO3, PO7, O1), and right posterior (CP2, CP4, CP6, TP8, P2, P4, P6, P8, PO4, PO8, O2). In all ANOVAs, all dependent variables were normally distributed and met the assumption of sphericity, unless otherwise indicated. The p-values of all pairwise-comparisons were Bonferroni corrected.

The visual inspection of the target word revealed an N400 component that is higher for false than for true affirmative sentences, but lower for false than for true negative sentences (see **Figure 1**). The ANOVA in the time-window 400–500 ms revealed an interaction *AP*Polarity*Truth* [F(1, 30) = 6.197, $p = 0.019$, partial $\eta^2 = 0.171$] as well as an interaction *Lateralization*Polarity*Truth* [F(1, 30) = 4.540, $p = 0.041$, partial $\eta^2 = 0.131$]. Furthermore, there is a main effect of *AP* [F(1, 30) = 21.599, $p = < 0.001$, partial $\eta^2 = 0.419$], with the frontal electrodes on average showing more negative amplitudes than the posterior electrodes ($\Delta_{(Post,Front)} = 1.962\,\mu V$), as well as a main effect of *Lateralization* [F(1, 30) = 16.136, $p = < 0.001$, partial $\eta^2 = 0.350$], with the electrodes on the right hemisphere showing more negative amplitudes than the left hemisphere ($\Delta_{(Right,Left)} = -0.863\,\mu V$). Subsequently, we performed ANOVAs for each region separately to break down the two interactions.

The separate ANOVA for the right posterior region revealed an interaction *Polarity*Truth* [F(1, 30) = 6.616, $p = 0.015$, partial $\eta^2 = 0.181$], with false affirmative sentences having more negative amplitudes than true affirmative sentences ($\Delta_{(False,True)} = -0.686\,\mu V$), and true negative sentences showing more negative amplitudes than false negative sentences ($\Delta_{(True,False)} = -0.386\,\mu V$). Broken down by *Polarity*, the separate ANOVAs revealed a main affect of *Truth* for the affirmative sentences [F(1, 30) = 7.453, $p = 0.01$, partial $\eta^2 = 0.199$], but not for the negative sentences [F(1, 30) = 1.979, $p = 0.170$, partial $\eta^2 = 0.062$]. See **Figure 2** for the topographical distribution. There is no main effect of *Polarity* ($F > 0.1$, $p > 0.7$) and no main effect of *Truth* ($F > 0.8, p > 0.3$).

No effects were found in the remaining three regions ($F > 0.03, p > 0.4$). Adding *Working Memory* or *AQ-score* as between-subject factors based on Median Split brought no significant effect for that factor[13].

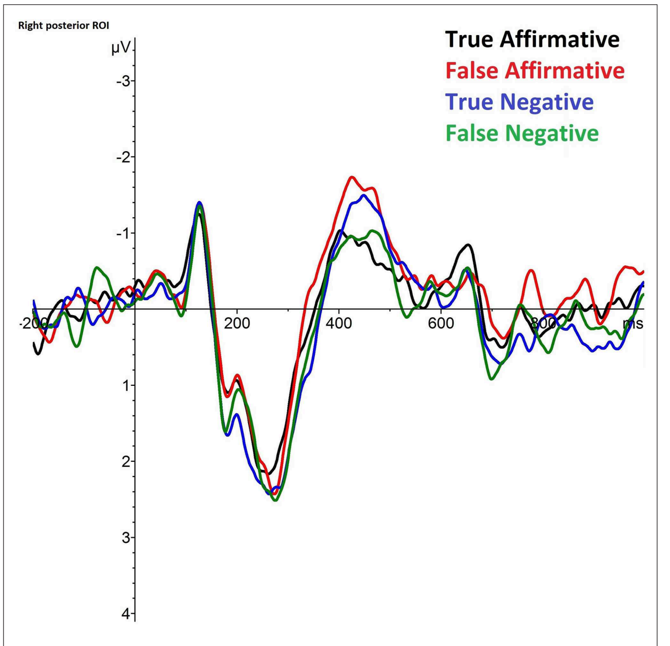


FIGURE 1 | Grand average across all subjects ($N = 31$) for all four experimental conditions at the right posterior ROI, based on the average of 11 electrode positions (CP2, CP4, CP6, TP8, P2, P4, P6, P8, PO4, PO8, O2).

The midline electrodes were analyzed separately with the factors *Polarity*, *Truth* and *Midline ROI* (anterior, posterior, left, right). The following electrodes are included: anterior (Fz, FCz, Cz), posterior (CPz, Pz, POz, Oz), left (C2, C3, C5, T7), right (C2, C4, C6, T8). The ANOVA revealed an interaction *Polarity*Truth*Midline ROI* [F(3, 28) = 2.884, $p = 0.04$, partial $\eta^2 = 0.088$], as well as a main effect for *Midline ROI* [$F(3, 28) = 15.348$, $p = < 0.001$, partial $\eta^2 = 0.338$]. Subsequently, we performed ANOVAs for each midline region separately to break down the interaction.

The posterior midline ROI shows a marginally significant interaction *Polarity*Truth* [F(1, 30) = 3.811, $p = 0.06$, partial $\eta^2 = 0.113$]. The right midline ROI shows a marginally significant interaction *Polarity*Truth* [F(1, 30) = 3.508, $p = 0.07$, partial $\eta^2 = 0.105$] as well. There was no main effect in any of the four midline regions, nor any interaction in the remaining two regions (anterior and left) ($F > 0.009, p > 0.4$).

3.3. Knowledge Questionnaire

In the first task of the post-experiment questionnaire participants had to rate each name used in the stimuli of the EEG experiment on a scale from 1–4[14] to indicate their level of knowledge. The second part of the questionnaire consisted of a sentence

[13] A figure with the Grand Averages split by Working Memory (high vs. low) as well as a figure of the separate Grand Averages for each of the four ROIs can be found in the **Supplementary Material**.

[14] The scale was identical to the pre-experiment questionnaire desribed in the section 2.2 (i.e., a four point Likert scale). Participants were informed in the following way about the scale: 4 = you know a person's name and profession; 3 = you know to whom the name refers to but have little knowledge about that person, e.g., you roughly know that somebody is from politics; 2 = you hardly know a person, that is, you heard the name before but do not know who that person is; 1 = you do not know a person at all. An example was used to demonstrate the distribution of the scale.

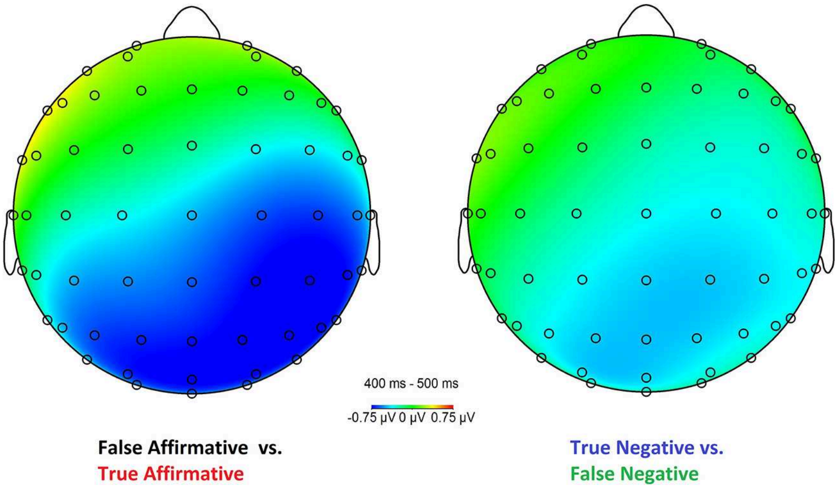


FIGURE 2 | Topographical maps of the differences between false and true affirmative sentences **(Left)** and between true and false negative sentences **(Right)** in the time window 400–500 ms. $N = 31$.

completion task asking them to complete *X war/ist* (X was/is), with X being the names used in the experiment. For the first task, the mean rating across all items and participants was 3.67 with SD = 0.82. Six subjects had mean values below 3, however, we did not exclude them from the analysis since their responses to the second part indicated that they had the required knowledge to assess the sentence truth-value. As indicated in the description of the material, 23 names that were not part of the pre-test were included in the stimuli. To assess our participants knowledge about these 23 items, we calculated the mean values for these 23 names separately, in addition to the analysis above. These 23 items received a mean of 3.05 with $SD = 0.96$. Among these 23 names, those items that received mean values below 3 interestingly received correct answers in all but two cases in the second task, indicating that our subjects had enough knowledge to recognize the truth of the stimuli sentences in most cases.

3.4. *Post-hoc* Questionnaire: Perceived Naturalness and Perceived Truth-Value Rankings

The *post-hoc* questionnaire was completed online by 80 participants who did not take part in the ERP-study. The mean ratings with standard deviations for perceived naturalness and perceived truth-values are shown in **Table 3**. Regarding the perceived naturalness, the ANOVA revealed a main effect of *Polarity* [F(1, 79) = 934.424, $p < 0.001$, partial $\eta^2 = 0.922$] and a main effect of *Truth* [F(1, 79) = 21.225, $p < 0.001$, partial $\eta^2 = 0.212$], as well as an interaction *Polarity*Truth* [F(1, 79) = 7.196, $p = 0.009$, partial $\eta^2 = 0.083$]. Broken down by *Polarity*, the separate ANOVAs revealed a main effect of *Truth* for the affirmative sentences [F(1, 79) = 17.212, $p < 0.001$, partial $\eta^2 = 0.179$], with lower mean ratings for false compared to true affirmative sentences ($\Delta_{(True,False)} = 0.22\,\mu V$). The negative sentences showed a main effect of *Truth* [F(1, 79) = 4.128, $p = 0.046$, partial $\eta^2 = 0.05$] as well, with slightly

TABLE 3 | Mean values of the perceived naturalness of the stimuli (scale: 4 = natural, 3 = rather natural, 2 = rather unnatural, 1 = unnatural) and mean values of the perceived truth (4 = true, 3 = rather true, 2 = rather false, 1 = false) for all four conditions; standard deviations are indicated in brackets.

	Naturalness		Truth	
	True	False	True	False
Affirmative	3.39 (3.42)	3.17 (3.5)	3.41 (6.2)	1.68 (5.48)
Negative	2.37 (1.71)	2.31 (1.92)	3.2 (5.79)	1.58 (5.44)

N = 80.

lower mean ratings for false compared to true negative sentences ($\Delta_{(True,False)} = 0.06\,\mu V$).

Regarding the perceived truth-value, the ANOVA revealed a main effect of *Polarity* [F(1, 79) = 5.464, $p = 0.022$, partial $\eta^2 = 0.65$] with affirmative sentences receiving higher mean ratings than negative sentences ($\Delta_{(True,False)} = 0.16$), and a main effect of *Truth* [$F(1, 79) = 876.789$, $p < 0.001$, partial $\eta^2 = 0.917$] with true sentences receiving higher mean values than false sentences ($\Delta_{(True,False)} = 1.67$). There was no interaction *Polarity*Truth* ($F > 0.593, p > 0.44$).

4. DISCUSSION

The comprehension of isolated negated sentences has been argued to be an exception to incremental language comprehension, which is based (inter alia) on evidence from a range of event-related potential studies showing an interaction of polarity and truth-value. These studies reported true negated sentences eliciting higher N400 ERPs than false negated sentences suggesting that the N400 is driven by priming relations within sentences rather than by sentence truth-value (Fischler et al., 1983; Wiswede et al., 2013; Dudschig et al., 2019). Our study examined the comprehension of negated sentences in comparison to affirmative sentences, employing world knowledge in true and false sentences.

In contrast to earlier studies that used similar designs, we make use of cohyponyms, thereby increasing the overlap of features from the set of true alternatives for the affirmative sentence and the set of true alternatives for the negative sentence. Thereby, we aimed at facilitating the anticipation in the negated sentences to investigate whether it leads to similar ERP-patterns as in previous studies (Fischler et al., 1983; Wiswede et al., 2013; Dudschig et al., 2019). *Anticipation* here is used to describe potential pre-activations of upcoming content in the linguistic input that results from a feature overlap with previously processed content. We hypothesized to find an N400 effect for false compared to true affirmative sentences, in line with earlier studies using world knowledge violations (Fischler et al., 1983; Hagoort et al., 2004; Nieuwland and Kuperberg, 2008; Wiswede et al., 2013; Metzner et al., 2015; Dudschig et al., 2016, 2019). For negative sentences, we hypothesized a reduction of the N400 effect for true vs. false sentences, or an N400 for false compared to true negative sentences.

We observe an interaction of truth-value and polarity which is driven by reversed effects for affirmative and negative sentences, that is, by larger N400 ERPs for false compared to true affirmative sentences, but smaller N400 ERPs for false compared to true negative sentences. Split by *Polarity*, the effect is significant only for affirmative sentences, with a larger N400 for false compared to true sentences. For negated sentences, there is no significant effect, but the trend goes in the same direction as in earlier studies, thus, negation reverses the N400 pattern with more negative amplitudes for true compared to false sentences. A significant interaction is observed in the right posterior region, matching the typical topography of the N400 component for written sentences (Kutas et al., 1988; Kutas and Federmeier, 2011). Overall, our amplitude differences seem to be smaller compared to earlier studies, which can be a result of the increased feature overlap between the alternatives used as critical words in our sentences. However, the decrease in amplitude size might as well be at least partially affected by the use of a probe task instead of a truth-value judgment task as in earlier studies which is in line with the results by Wiswede et al. (2013), who observed reduced amplitude differences in the N400 time window for the control group compared to the truth-evaluation group.

The ERP-results match the observations for reaction times to the probe task which show longer response times for false compared to true affirmative sentences, but no difference between negative sentences. Additionally, there is no significant difference between responses to affirmative and negative sentences, yet the means show that across conditions, responses to the false affirmative sentences were the slowest.

Furthermore, we assumed that engaging in anticipatory mechanisms can be expected to be more difficult for individuals with lower working memory capacities (WMC), which eventually may lead to an absence of such mechanisms in this group. Therefore, we hypothesized that participants with low WMC will show lower N400 effects than people with high WMC since the latter may anticipate a range of alternatives more easily, resulting in stronger N400s for each of them. While the visual inspection of the data shows that the N400 amplitudes are generally reduced for people with low working memory capacities, the correlations with the working memory tests (Reading Span and Digit Span) were not significant. The results partially match the findings from Otten and van Berkum (2009) who investigated the impact of individual WMC in an ERP-study. They report individuals with low as well as with high working memory capacities to predict specific upcoming words. Both groups show an early negative deflection for unexpected compared to expected determiners in predictive stories. Hence, the ability to rapidly and automatically predict upcoming linguistic material seems to be independent of a person's WMC, that is, of their ability to temporarily store and manipulate information. At the same time, however, in the study by Otten and van Berkum (2009), low working memory readers additionally showed a late negativity to linguistic material that was inconsistent with the participant's prediction, suggesting additional processing. Possibly, this additional neural response reflects increased demands of the adjustment or the suppression of the original prediction. While this result matches the findings of Luedtke et al. (2008), who report an enhanced negativity for words following the negative quantifier *no* (e.g., *In front of the tower there is no ghost*), it does not match the results from the current study. Since we did not select participants based on their working memory capacities and since our participants are mostly students in a certain age range, the variation of values they obtained in the different pre-tests are mainly pooled at the upper and upper-central part of the respective scales. Therefore, the variation resulting from grouping them into high-working-memory-readers and low-working-memory-readers based on a median split might have been too low to become significant, especially since the N400 amplitude differences in our study are generally reduced. We were furthermore interested to see whether the N400-ERPs for negated sentences, which are typically underinformative, at least when presented in isolation, correlate with the AQ-score of participants, similarly to Nieuwland et al. (2010). We do not find such a correlation, matching the results by Spychalska et al. (2016). Again, we did not select participants based on their scores in the AQ-test and neurological and psychological disorders were an exclusion criterion for our study.

One potential explanation for larger effects within the affirmative sentences is the anticipation of upcoming content in the true affirmative compared to the false affirmative sentences due to a higher overlap of features associated with e.g., *Angelina Jolie* and *actress* than with *singer*. In negative sentences, instead, the anticipation of alternatives is usually more difficult, unless the context provides only a binary choice of alternative options (Orenes et al., 2014). The reduction of the amplitude difference between the two negated conditions might reflect a "partially incremental" integration of the negation which was facilitated due to the feature overlap within both sentences. A fully incremental integration should have led to an N400 for false compared to true negative sentences. Instead, a total absence of incremental comprehension should have led to the same amplitude differences as for affirmative sentences. Urbach and Kutas (2010) provide a similar suggestion, namely, that the interpretation of quantifier expressions as for example *most* and *few* is neither fully incremental nor fully delayed, therefore, it is argued to be "partially incremental." Sentences with negative quantifiers have been reported to reveal similar result patterns as sentences with propositional negation, resulting in an interaction of truth and quantifier type (Kounios and Holcomb, 1992). Related to that, Nieuwland (2016) observed smaller N400s for false compared to true sentences, independent of the type of quantifier (*few* vs. *many*), however only in sentences with high cloze values for the target word. For sentences with lower cloze values, the pattern for positive quantifiers was similar, but it was reversed for negative quantifiers, that is, in sentences where the target word had a low cloze value, the true negative sentences had higher N400 amplitudes than the false negative sentences. Even though our affirmative and negative sentences are not matched with regard to cloze value, the increased feature overlap in our study might facilitate the anticipation of upcoming content in a similar way. The smaller difference between amplitudes for our negative sentences might therefore reflect an approximation toward a typical N400 pattern, with false compared to true sentences eliciting larger N400s. Based on our results and the results by Nieuwland (2016), high cloze values then should further affect the N400 for sentences with propositional negation, leading to a similar pattern as is typically observed for affirmative sentences, that is, a larger N400 for false over true sentences.

Negation has been reported to lead to lower activation levels for negated probes (MacDonald and Just, 1989) and negated sentences (Tettamanti et al., 2008) in functional neuroimaging studies. As part of the related debate about negation playing some sort of inhibitory role on concepts, it has been discussed whether this attenuation also spreads to associated concepts or whether the negation of one concept actually enhances a spread of activation across associated alternative concepts (see e.g., Anderson et al., 2010). Given that negations tend to occur in the same contexts as their affirmative alternatives, the latter option seems to support an incremental and efficient comprehension process more than the former. Furthermore, without alternatives, negated sentences would be underinformative. MacDonald and Just (1989) did not find an inhibitory effect of negation on associated concepts, suggesting that those alternatives indeed became activated during the comprehension process, which matches our results. The use of alternatives furthermore depends on the negated dimension, hence, on the scope of the negation. Previous studies suggested mixed results about alternatives during comprehension. Tian and Breheny (2015) have shown that in sentences with clear scope and therefore clearer alternatives (e.g., *It is John who hasn't ironed his brother's shirt*), incremental comprehension is facilitated. However, reducing the alternatives alone does not facilitate comprehension in all cases. Nieuwland and Kuperberg (2008) used contrary adjectives such as e.g., *easy-difficult, rich-poor, safe-dangerous* as target words which, when negated, directly allow for an interpretation by replacing the negated adjective with its unique alternative. Yet, this alone did not facilitate comprehension, instead, only the pragmatic embedding of the sentences did. In our sentences, having wide scope, in principle everything could have been negated, including both nouns, the verb as well as the prepositional phrase. However, negating the first noun would require further emphasis, either by stressing it (in spoken language) or by using a cleft sentence, e.g., *It is not George Clooney who is an actor*. In theory, in our material it is the second noun and the prepositional phrase that can be interpreted as being negated either altogether or separately. We cannot exclude that participants interpreted the negation taking scope over the final phrase of our sentences. However, it is likely that they noticed that it is the profession (second noun) rendering some of the sentences true and others false because the final phrase always led to true sentences across all stimuli and across fillers.

Due to the addition of the final phrase, e.g., *in the USA*, that was added to avoid a potential overlap of negation-induced effects and the sentence wrap-up effect (Just and Carpenter, 1980; Rayner et al., 2000; Hirotani et al., 2006; Warren et al., 2009), one might argue that participants could have "waited" for this phrase to come for their intuitive truth-value judgment. However, first of all, no explicit truth-value judgment was required and under the notion of incremental comprehension the target word can be assumed to be integrated into sentence meaning before the occurrence of further input material. Secondly, the final phrase did not modulate the truth-value, but provided the true origin of the person mentioned in the sentences. Yet, intuitive truth-value judgment, despite not being required for the task, was required to achieve a full understanding of the sentence which might have been especially difficult for the true negative sentences. For example, *George Clooney is not a singer* might be intuitively easy to be judged as true. Yet, to fully assess its truth, we usually do not have enough knowledge about celebrities. As a result, participants might have achieved a full understanding only for the affirmative sentences which clearly show an N400 effect despite the implicit probe task, but not for the negated sentences. The second aspect of our design that might have had further impact on the differences across conditions is the adverb in the affirmative conditions. To keep the sentence length stable, the adverbs *derzeit* ("currently") or *einmal* ("once") were added into the affirmative sentences. They were chosen as the best fit under the additional constraint of having similar frequency values as the negative operator[15]. We cannot exclude that the use of these adverbs had some effect on the results, leading to the stronger N400 contrast for the affirmative sentences. While the adverb *derzeit* ("currently") seems a relatively neutral counterpart to a negation for the sentences with present tense, the adverb *einmal* ("once") might have had a pragmatic effect on the interpretation. While it can be understood along the lines of *used to be* when combined with the verb *to be*, some people might also interpret it more strictly as meaning *one time* which would make the interpretation of the false affirmative sentences more difficult because we do not have enough knowledge about the people described by the sentence, to exclude the possibility that, for example, *Beethoven once was a painter* is false because we might assume that maybe he indeed also painted and we simply do not know about it.

[15]Nevertheless, we would like to point out that negative marker "nicht" is a very high frequent word (ranked 17th of all words in the database) which inevitably leads to rather strong frequency differences when compared to any adjective both in the current as well as in previous studies comparing affirmative and negative sentences.

The typically reversed effect for negated sentences with the N400 being larger for true compared to false negative sentences has often been assumed to be driven by the true negative sentences and the semantic distance due to lower feature overlap between the entities within these sentences (e.g., *rose* and *insect* Fischler et al., 1983 or *George Clooney* and *singer* in our study) and therefore, the absence of priming when compared to the false negative sentences. Alternatively, the reversed ERP-pattern might be driven by the falsity of the false negative sentences not being detected. Incomplete or "shallow" processing in general refers to an incomplete interpretation of the information available in the linguistic input which results in an incomplete or underspecified representation (Frazier and Rayner, 1990; Ferreira et al., 2002; Sanford and Sturt, 2002, see also Baggio et al., 2012). It has been observed in various experiments that in certain scenarios where the semantic similarity between words of a sentence is high, readers do not detect incomplete or semantically anomalous information, thereby achieving a wrong interpretation of the sentence. Examples for these kind of "semantic illusions" are "How many animals of each type did Moses take on the ark?" (Erickson and Mattson, 1981) where readers did not detect that it was not Moses but Noah who took animals onto the arch, or "What is the holiday where children go door to door, dressed in costumes, giving out candy?" (Reder and Kusbit, 1991), where participants fail to detect that children do not hand out but collect candies. Potentially, in our study, participants failed to detect the falsity of sentences like *George Clooney is not an actor* or, in the study by Fischler et al. (1983) of *A bee is not an insect*, at least not fast enough for the difference to be reflected in the online-comprehension signature. Even though this is a null result, we point out that reaction time data further support this interpretation, as there was no difference between the responses to the negated sentences, but there was a difference between the affirmative sentences. This potential interpretation matches the results from the quantifiers study by Urbach and Kutas (2010) mentioned before, who tested sentences like *Most/Few farmers grow crops/worms* and reported an N400 for *worms* independent of the quantifier type. However, the effect was smaller for cases with negative quantifiers. At the same time, the offline plausibility judgment showed that both true sentences were rated more plausible suggesting that participants achieved a full understanding by the time the plausibility question appeared after the sentence. It has been shown in earlier studies, that manipulating the time window between an affirmative or negative sentence and a subsequently presented matching or mismatching picture leads to different results suggesting that the implementation of the negation into the sentence meaning is time consuming (Kaup et al., 2006; Luedtke et al., 2008). Our ERP-results then would not contradict those findings. At the same time, however, they suggest that an increase of feature overlap between the entities of a sentence seems to trigger a "partially incremental" interpretation of the negation operator.

The ratings of our *post-hoc* questionnaire, in which participants were asked to rate all sentences regarding their naturalness and their truth-values, each on a scale from 1 to 4, suggest that negative sentences are in general perceived to be less natural than affirmative sentences. This finding is not surprising as negative sentences are less frequent and more marked, and especially when presented in isolation are considered to be less informative than affirmative sentences, thereby violating Grice's Conversational Maxims (Grice, 1975). Furthermore, false sentences were perceived as slightly less natural than true sentences. Therefore, the reversed N400 for true compared to false negative sentences cannot be explained based on their perceived naturalness. If naturalness was the reason for the observed N400 in this and prior studies, true negative sentences should have received lower mean values with respect to their naturalness than false negative sentences. Regarding the perceived truth-value, participants responses are as expected, that is, true sentences received higher ratings than false sentences, independent of their polarity. Furthermore, affirmative sentences received higher ratings than negative sentences. It should be noted though that the standard deviations are relatively high for all four conditions, indicating that individual responses strongly varied. These results match the accuracy of responses to the main experiment with false negative sentences leading to more incorrect responses both compared to true negative as well as compared to false affirmative sentences. Given that this experiment involves world knowledge and given the addition of the adverb as well as the final phrase along with their pragmatic implications, the results of the *post-hoc* questionnaire are not surprising. Importantly though, as shown by the main effect of *Truth*, false sentences can be expected to be recognized as such, both for affirmative as well as for negative sentences. However, note that the questionnaire ratings are offline-ratings. Participants were asked to rely on intuitive judgments without thinking too long about each sentence. Yet, the presentation time of the sentences was not limited in time and furthermore, sentences were not split into single words. Instead, in the ERP-study, which reflects online-sentence comprehension, each word was presented in isolation and only for few hundred milliseconds. Taken together, the pattern from our ERP-study and from our *post-hoc* questionnaire ratings fully match the combination of online-ERP results and offline plausibility judgments by Urbach and Kutas (2010) presented above. Furthermore, one might argue that the truth-value ratings contradict the claim made above that the falsity of the false negated sentences is not detected. However, let us emphasize again that the option of the falsity of the negative sentence not being detected fast enough, hence, at the time the target word is presented, cannot be ruled out by the questionnaire. Hence, the offline questionnaire rating does not contradict the option of an incomplete computation of the negated sentence meaning that is only completed later in time.

In sum, our study can be taken as an indication for an increase of feature overlap between the entities within sentences leading to a decrease of amplitude differences between true and false negative sentences compared to earlier studies, however, the trend for negative sentences eliciting larger N400s than false negative sentences persists. Our results are in line with earlier studies, but additionally they suggest a "partially incremental" comprehension process, that is, the integration of the negation with the information in its scope is neither fully incremental nor fully delayed. Future experiments investigating the time-course of comprehension in negated sentences using different verb types and varying the position of the negative marker regarding the verb are necessary, for example to assess the role of alternatives along other dimensions as well as the role of alternatives in sentences with full verbs opposed to copula verbs in general. In addition, comparing affirmative and negative sentences with similar cloze values (cf. Nieuwland, 2016) could reveal further information about (isolated) negated sentences being processed incrementally or not.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/**Supplementary Files**.

ETHICS STATEMENT

The study was conducted in compliance with the Helsinki declaration and the ethical guidelines of the national research funding agency in Germany (German Research Foundation, DFG). The presented material did not contain any potentially harmful content and EEG is a non-invasive method. According to the guidelines of the German Research Foundation, for EEG experiments, an explicit ethics approval is only required if the study is conducted with persons below 18 or over 65 years old, or with patients. All subjects were adult volunteers (between 18 and 38 years old) and signed an informed consent of participation, which included information about protection of privacy and confidentiality, as well as the right to withdraw from participation at any point.

AUTHOR CONTRIBUTIONS

VH designed, programmed and conducted the experiment, prepared the stimuli, processed and analyzed the data, and wrote the manuscript. MS advised on the data processing, interpretation as well as and on the manuscript, and contributed to the data analysis. MW advised on the experiment design, the data processing, data interpretation, and provided the resources for the experiment.

FUNDING

We are grateful to the Foundation Stiftung Mercator for financial support through the Research Group Structure of Memory as well as to the German Research Foundation (DFG) for financial support through the grants WE 4984/4-1 and SP 1718/1-1 as part of the Priority Program XPrag.de (SPP1727). We acknowledge support by the DFG Open Access Publication Funds of the Ruhr-Universität Bochum.

ACKNOWLEDGMENTS

We would like to thank our student assistants for their support with the data collection. We thank Erica Cosentino, Regina Fabry, Maria Piñango, and Matthias Unterhuber for helpful discussions, and Daniel Labbé for assistance with the programming. We thank both reviewers as well as the editors for their helpful suggestions and comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02184/full#supplementary-material

REFERENCES

Anderson, S., Huette, S., Matlock, T., and Spivey, M. (2010). "On the temporal dynamics of negated perceptual simulations," in *Meaning Form and Body*, eds F. Parrill, V. Tobin, and M. Turner (Stanford, CA: CSLI Publications, Stanford University), 1–20.

Baggio, G., van Lambalgen, M., and Haagort, P. (2012). "The processing consequences of compositionality," in *The Oxford Handbook of Compositionality*, eds W. Hinzen, E. Machery, and M. Werning (Oxford: Oxford University Press), 657–674.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471

Borovsky, A., Elman, J. L., and Fernald, A. (2012). Knowing a lot for one's age: vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *J. Exp. Child Psychol.* 112. 417–436. doi: 10.1016/j.jecp.2012.01.005

Carpenter, P. A., and Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychol. Rev.* 82, 45–73. doi: 10.1037/h0076248

Chow, W.-Y., Lau, E., Wang, S., and Phillips, C. (2018). Wait a second! delayed impact of argument roles on on-line verb prediction. *Lang. Cogn. Neurosci.* 33, 803–828. doi: 10.1080/23273798.2018.1427878

Clark, H. H., and Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cogn. Psychol.* 3, 472–517. doi: 10.1016/0010-0285(72)90019-9

Cosentino, E., Baggio, G., Kontinen, J., and Werning, M. (2017). The time-course of sentence meaning composition. N400 effects of the interaction between context-induced and lexically stored affordances. *Front. Psychol.* 8:813. doi: 10.3389/fpsyg.2017.00813

Dale, R., and Duran, N. (2011). The cognitive dynamics of negated sentence verification. *Cogn. Sci.* 35, 983–996. doi: 10.1111/j.1551-6709.2010.01164.x

Dambacher, M., Kliegl, R., Hoffmann, M., and Jacobs, A. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Res.* 1084, 89–103. doi: 10.1016/j.brainres.2006.02.010

De Long, K., Urbach, T., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504

DeLong, K. A., Groppe, D. M., Urbach, T. P., and Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain Lang.* 121, 226–239. doi: 10.1016/j.bandl.2012.02.006

Dudschig, C., Mackenzie, I. G., Maienborn, C., Kaup, B., and Leuthold, H. (2019). Negation and the N400: investigating temporal aspects of negation integration using semantic and world-knowledge violations. *Lang. Cogn. Neurosci.* 34, 309–319. doi: 10.1080/23273798.2018.1535127

Dudschig, C., Maienborn, C., and Kaup, B. (2016). Is there a difference between stripy journeys and stripy ladybirds? The n400 response to semantic and world-knowledge violations during sentence processing. *Brain Cogn.* 103, 38–49. doi: 10.1016/j.bandc.2016.01.001

Erickson, T. D., and Mattson, M. E. (1981). From words to meaning: a semantic illusion. *J. Verb. Learn. Verb. Behav.* 20, 540–551. doi: 10.1016/S0022-5371(81)90165-1

Federmeier, K. D., and Kutas, M. . (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. doi: 10.1006/jmla.1999.2660

Federmeier, K. D., and Kutas, M. (2005). Aging in context: age-related changes in context use during language comprehension. *Psychophysiology* 42, 133–141. doi: 10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., McLennan, D. B., De Ochoa, E., and Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERB study. *Psychophysiology* 39, 133–146. doi: 10.1111/1469-8986.3920133

Ferguson, H. J., Sanford, A. J., and Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Res.* 1236, 113–125. doi: 10.1016/j.brainres.2008.07.099

Ferreira, F., Ferraro, V., and Baile, K. (2002). Good-enough representations in language comprehension. *Curr. Dir. Psychol. Sci.* 11, 11–15. doi: 10.1111/1467-8721.00158

Fischler, I., Bloom, P., Childers, D., Roucos, S., and Perry, N. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology* 20, 400–409. doi: 10.1111/j.1469-8986.1983.tb00920.x

Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain Lang.* 140, 1–11. doi: 10.1016/j.bandl.2014.10.006nguage

Frazier, L., and Rayner, K. (1990). Taking on semantic commitments: processing multiple meanings vs. multiple senses. *J. Mem. Lang.* 29, 181–200. doi: 10.1016/0749-596X(90)90071-7

Grice, H. P. (1975). "Logic and conversation," in *Syntax and Semantics*, Vol. 3, eds P. Cole and J. Morgan (New York, NY: Academic Press), 41–58. Reprinted in *Studies in the Way of Words*.

Gunter, T., Friederici, A., and Hahne, A. (1999). Brain responses during sentence reading: visual input affects central processes. *NeuroReport* 10, 3175–3178. doi: 10.1097/00001756-199910190-00009

Gunter, T., Stowe, L., and Mulder, G. (1997). When syntax meets semantics. *Psychophysiology* 34, 660–676. doi: 10.1111/j.1469-8986.1997.tb02142.x

Hagoort, P., Hald, L. A., Bastiaansen, M. C. M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441. doi: 10.1126/science.1095455

Hasson, U., and Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *J. Pragmat.* 38, 1015–1032. doi: 10.1016/j.pragma.2005.12.005

Hirotani, M., Frazier, L., and Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: evidence from eye movements. *J. Mem. Lang.* 54, 425–443. doi: 10.1016/j.jml.2005.12.001

Ito, A., Martin, A. E., and Nieuwland, M. S. (2016). How robust are prediction effects in language comprehension? Failure to replicate article-elicited n400 effects. *Lang. Cogn. Neurosci.* 32, 954–965. doi: 10.1080/23273798.2016.1242761

Ito, A., Martin, A. E., and Nieuwland, M. S. (2017). Why the a/AN prediction effect may be hard to replicate: a rebuttal to delong, urbach, and kutas (2017). *Lang. Cogn. Neurosci.* 32, 974–983. doi: 10.1080/23273798.2017.1323112

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354. doi: 10.1037/0033-295X.87.4.329

Kaup, B., Luedtke, J., and Zwaan, R. (2007). "The experiental view of language comprehension: How is negated text information represented?," in *Higher Level Language Processes in the Brain: Inference and Comprehension Processes*, eds F. Schmalhofer and C. Perfetti (Mahwah, NJ: Lawrence Erlbaum), 255–288.

Kaup, B., Luedtke, J., and Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: is a door that is not open mentally closed? *J. Pragmat.* 38, 1033–1050. doi: 10.1016/j.pragma.2005.09.012

Kay, P. (1971). Taxonomy and semantic contrast. *Language* 47:866. doi: 10.2307/412161

Kounios, J., and Holcomb, P. J. (1992). Structure and process in semantic memory: evidence from event-related brain potentials and reaction times. *Can. J. Exp. Psychol.* 121, 459–479. doi: 10.1037/0096-3445.121.4.459

Kruszewski, G., Paperno, D., Bernardi, R., and Baroni, M. (2016). There is no logical negation here, but there are alternatives: modeling conversational negation with distributional semantics. *Comput. Linguist.* 42, 637–660. doi: 10.1162/COLI_a_00262

Kuperberg, G. (2010). Language in schizophrenia part 1: an introduction. *Linguist. Lang. Compass* 4, 576–589. doi: 10.1111/j.1749-818X.2010.00216.x

Kuperberg, G., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299

Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/S1364-6613(00)01560-6

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0

Kutas, M., van Petten, C., and Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalogr. Clin. Neurophysiol.* 69, 218–233. doi: 10.1016/0013-4694(88)90131-9

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discour. Process.* 25, 259–284. doi: 10.1080/01638539809545028

Lau, E., Phillips, C., and Poeppel, D. (2009). A cortical network for semantics: (de)constructing the n400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Luedtke, J., Friedrich, C. K., De Filippi, M., and Kaup, B. (2008). Event-related potential correlates of negation in a sentence-picture verification paradigm. *J. Cogn. Neurosci.* 20, 1355–1370. doi: 10.1162/jocn.2008.20093

MacDonald, M., and Just, M. A. (1989). Changes in activation levels with negation. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 633–642. doi: 10.1037/0278-7393.15.4.633

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., and Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *J. Mem. Lang.* 69, 574–588. doi: 10.1016/j.jml.2013.08.001

Metzner, P., von der Malsburg, T., Vasishth, S., and Roesler, F. (2015). Brain responses to world knowledge violations: a comparison of stimulus- and fixation-triggered event-related potentials and neural oscillations. *J. Cogn. Neurosci.* 27, 1017–1028. doi: 10.1162/jocn_a_00731

Mohammad, S. M., Dorr, B. J., Hirst, G., and Turney, P. D. (2013). Computing lexical contrast. *Comput. Linguist.* 39, 555–590. doi: 10.1162/COLI_a_00143

Nieuwland, M. S. (2016). Quantification, prediction, and the online impact of sentence truth-value: evidence from event-related potentials. *J. Exp. Psychol.* 42, 316–334. doi: 10.1037/xlm0000173

Nieuwland, M. S., Ditman, T., and Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: an ERP investigation of informativeness and pragmatic abilities. *J. Mem. Lang.* 63, 324–346. doi: 10.1016/j.jml.2010.06.005

Nieuwland, M. S., and Kuperberg, G. R. (2008). When the truth isn't too hard to handle: an event-related potential study on the pragmatics of negation. *Psychol. Sci.* 19, 1213–1218. doi: 10.1111/j.1467-9280.2008.02226.x

Nieuwland, M. S., and Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *J. Cogn. Neurosci.* 18, 1098–1111. doi: 10.1162/jocn.2006.18.7.1098

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Orenes, I., Beltran, D., and Santamaria, C. (2014). How negation is understood: evidence from the visual world paradigm. *J. Mem. Lang.* 74, 36–45. doi: 10.1016/j.jml.2014.04.001

Otten, M., and van Berkum, J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Res.* 1291, 92–101. doi: 10.1016/j.brainres.2009.07.042

Rayner, K., Kambe, G., and Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *Q. J. Exp. Psychol. Sect. A* 53, 1061–1080. doi: 10.1080/713755934

Reder, L. M., and Kusbit, G. W. (1991). Locus of the moses illusion: imperfect encoding, retrieval, or match? *J. Mem. Lang.* 30, 385–406. doi: 10.1016/0749-596X(91)90013-A

Rosch, E. (1975). Cognitive representations of semantic categories. *J. Exp. Psychol.* 104, 192–233. doi: 10.1037/0096-3445.104.3.192

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cogn. Psychol.* 8, 382–439. doi: 10.1016/0010-0285(76)90013-X

Rosch, E. H. (1973). Natural categories. *Cogn. Psychol.* 4, 328–350. doi: 10.1016/0010-0285(73)90017-0

Sanford, A., and Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends Cogn. Sci.* 6:382. doi: 10.1016/S1364-6613(02)01958-7

Spychalska, M., Kontinen, J., Noveck, I., Reimer, L., and Werning, M. (2019). When numbers are not exact: ambiguity and prediction in the processing of sentences with bare numerals. *J. Exp. Psychol.* 45, 1177–1204. doi: 10.1037/xlm0000644

Spychalska, M., Kontinen, J., and Werning, M. (2016). Investigating scalar implicatures in a truth-value judgment task: evidence from event-related brain potentials. *Lang. Cogn. Neurosci*. 31, 817–840. doi: 10.1080/23273798.2016.1161806

Tettamanti, M., Manenti, R., della Rosa, P., Falini, A., Perani, D., Cappa, S., et al. (2008). Negation in the brain: modulating action representations. *Neuroimage* 43, 358–367. doi: 10.1016/j.neuroimage.2008.08.004

Tian, Y., and Breheny, R. (2015). "Dynamic pragmatic view of negation processing," in *Negation and Polarity: Experimental Perspectives*, eds P. Larrivée and C. Lee (Cham: Springer International Publishing Switzerland), 21–43.

Tian, Y., Breheny, R., and Ferguson, H. (2010). Why we simulate negated information: a dynamic pragmatic account. *Q. J. Exp. Psychol.* 63, 2305–2312. doi: 10.1080/17470218.2010.525712

Urbach, T. P., and Kutas, M. (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *J. Mem. Lang.* 63, 158–179. doi: 10.1016/j.jml.2010.03.008

Van Berkum, J. J. A., Brown, C. M., and Hagoort, P. (1999). Early referential context effects in sentence processing: evidence from event-related brain potentials. *J. Mem. Lang.* 41, 147–182. doi: 10.1006/jmla.1999.2641

van den Noort, M., Bosch, P., Haverkort, M., and Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *Eur. J. Psychol. Assess.* 24, 35–42. doi: 10.1027/1015-5759.24.1.35

van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cogn.* 18, 380–393. doi: 10.3758/BF03197127

Warren, T., White, S. J., and Reichle, E. D. (2009). Investigating the causes of wrap-up effects: evidence from eye movements and e–z reader. *Cognition* 111, 132–137. doi: 10.1016/j.cognition.2008.12.011

Weckerly, J., and Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology* 36, 559–570. doi: 10.1111/1469-8986.3650559

Werning, M., Unterhuber, M., and Wiedemann, G. (2019). "Bayesian pragmatics provides the best quantitative model of context effects on word meaning in EEG and Cloze data," in *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, eds A. Goel, C. Seifert, and C. Freska (Austin, TX: Cognitive Science Society), 3085–3091.

Wiswede, D., Koranyi, N., Müller, F., Langner, O., and Rothermund, K. (2013). Validating the truth of propositions: behavioral and ERP indicators of truth evaluation processes. *Soc. Cogn. Affect. Neurosci.* 8, 647–653. doi: 10.1093/scan/nss042

Wlotko, E. W., and Federmeier, K. D. (2012). So that's what you meant! event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage* 62, 356–366. doi: 10.1016/j.neuroimage.2012.04.054

Wlotko, E. W., Federmeier, K. D., and Kutas, M. (2012). To predict or not to predict: age-related differences in the use of sentential context. *Psychol. Aging* 27, 975–988. doi: 10.1037/a0029206

Wolff, S., Schlesewsky, M., Hirotani, M., and Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: electrophysiological evidence from Japanese. *Brain Lang.* 107, 133–157. doi: 10.1016/j.bandl.2008.06.003

Yurchenko, A., den Ouden, D.-B., Hoeksema, J., Dragoy, O., Hoeks, J. C., and Stowe, L. A. (2013). Processing polarity: ERP evidence for differences between positive and negative polarity. *Neuropsychologia* 51, 132–141. doi: 10.1016/j.neuropsychologia.2012.10.028

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 October 2019
doi: 10.3389/fpsyg.2019.02214

Context, Content, and the Occasional Costs of Implicature Computation

*Raj Singh**

Institute of Cognitive Science, Carleton University, Ottawa, ON, Canada

The computation of scalar implicatures is sometimes costly relative to basic meanings. Among the costly computations are those that involve strengthening "some" to "not all" and strengthening inclusive disjunction to exclusive disjunction. The opposite is true for some other cases of strengthening, where the strengthened meaning is *less* costly than its corresponding basic meaning. These include conjunctive strengthenings of disjunctive sentences (e.g., free-choice inferences) and exactly-readings of numerals. Assuming that these are indeed all instances of strengthening via implicature/exhaustification, the puzzle is to explain why strengthening sometimes increases costs while at other times it decreases costs. I develop a theory of processing costs that makes no reference to the strengthening mechanism or to other aspects of the derivation of the sentence's form/meaning. Instead, costs are determined by domain-general considerations of the grammar's output, and in particular by aspects of the *meanings* of ambiguous sentences and particular ways they update the context. Specifically, I propose that when the hearer has to disambiguate between a sentence's basic and strengthened meaning, the processing cost of any particular choice is a function of (i) a measure of the semantic complexity of the chosen meaning and (ii) a measure of how much relevant uncertainty it leaves behind in the context. I measure semantic complexity with Boolean Complexity in the propositional case and with semantic automata in the quantificational case, both of which give a domain-general measure of the minimal representational complexity needed to express the given meaning. I measure relevant uncertainty with the information-theoretic notion of entropy; this domain-general measure formalizes how 'far' the meaning is from giving a complete answer to the question under discussion, and hence gives an indication of how much representational complexity is yet to come. Processing costs thus follow from domain-general considerations of current and anticipated representational complexity. The results might also speak to functional motivations for having strengthening mechanisms in the first place. Specifically, exhaustification allows language users to use simpler forms than would be available without it to both resolve relevant uncertainties and convey complex meanings.

Keywords: implicature, exhaustivity, complexity, processing, questions and answers, ambiguity

OPEN ACCESS

Edited by:
Katharina Spalek,
Humboldt University of Berlin, Germany

Reviewed by:
Emmanuel Chemla,
UMR8554 Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP), France
Andreas Haida,
Hebrew University of Jerusalem, Israel

****Correspondence:***
Raj Singh
raj.singh@carleton.ca

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *23 February 2019*
Accepted: *17 September 2019*
Published: *25 October 2019*

Citation:
Singh R (2019) Context, Content, and the Occasional Costs of Implicature Computation. Front. Psychol. 10:2214. doi: 10.3389/fpsyg.2019.02214

1. INTRODUCTION

1.1. Basic and Strengthened Meanings

It is commonly assumed that the 'basic meaning' of the sentence in (1)—the meaning as compositionally derived using the lexical items overtly present in the sentence—is the existential meaning ∃ in (1-a) that we learn in introductory logic. The sentence can of course be used to convey that Jan did not eat all of the cookies, ¬∀. This is not entailed by the sentence's basic meaning. Instead, the inference is commonly assumed to be an inference called the 'scalar implicature' of ∃ (1-b). Scalar implicatures are computed by a general mechanism that reasons about alternative propositions the speaker could have expressed but chose not to (in this case that Jan ate all of the cookies). The conjunction of (1)'s basic meaning with its scalar implicature is its "strengthened meaning" (1-c).

(1) Jan ate some of the cookies
  a. Basic meaning: that Jan ate some, possibly all, of the cookies (= ∃)
  b. Scalar implicature: that Jan did not eat all of the cookies (= ¬∀)
  c. Strengthened meaning: that Jan ate some but not all of the cookies (= ∃ ∧ ¬∀).

There is debate about the mechanism responsible for strengthening. For example, there are questions about whether the mechanism is part of the linguistic system itself or is shorthand for pragmatic or central-system reasoning. Putting this architectural question aside for the moment, all agree that the mechanism is an alternative-sensitive computation. More precisely, it is commonly assumed that there is a function, *STR*, which computes strengthened meanings by conjoining the sentence *S* with the negation of some of the alternatives of *S*, $ALT(S)$[1]. In general, *STR* is thought to be sensitive to various contextual factors, such as what is relevant, what is salient, what is assumed about the speaker's epistemic state, and other factors that have been identified in the literature. Thus, *STR* is a function that takes at least three inputs: the sentence *S*, its alternatives $ALT(S)$, and the context *c*, and returns the strengthened meaning of *S* in *c*, S_c^+: $STR(S, ALT(S), c) = S_c^+$. Thus, in a context *c* in which ∀ is relevant and the speaker is assumed to be opinionated about whether ∀ is true, $STR(\exists, ALT(\exists), c) = \exists_c^+ = \exists \wedge \neg\forall$. Suppose, however, that ∃ is uttered in a context c' in which ∀ isn't even relevant. In such a case, we say that the context has "pruned" ∀ from $ALT(\exists)$ (for more on pruning and constraints on pruning, see e.g., Magri, 2009; Fox and Katzir, 2011; Katzir, 2014; Crnič et al., 2015; Singh et al., 2016b). This pruning means that there are no alternatives left in *ALT* to negate, and hence application of *STR* would have no effect: $STR(\exists, ALT(\exists), c') = \exists_{c'}^+ = \exists$. In what follows, unless otherwise noted I will assume that we are in contexts in which all the members of *ALT* are relevant and that the speaker is opinionated about them. I will also sometimes disregard the distinction between a sentence and its denotation when there is little risk of confusion.

Competence theories of implicature computation need to specify *STR* and *ALT* and their interactions with the context such that the right strengthened meaning is derived for any sentence *S* in any context *c*. I will not spend much time discussing competing theories of these components. My concern in this paper is with exploring how competence-theoretic assumptions about strengthening might be realized in performance (see Chomsky, 1965 on the competence-performance distinction, and see Chemla and Singh (2014a,b) for the connection to experimental work on scalar implicature). As we will see, my strategy is to focus on the *output* of strengthening, not on the way in which strengthened meanings are actually derived. Specifically, I will explore the hypothesis that the processing costs that are sometimes associated with strengthening are derived entirely from considerations of the *meaning* of the sentence and specific ways in which it updates the context. The computational history of the sentence and its meaning will be irrelevant.

Nevertheless, to fix ideas it will be useful to assume a particular competence-theoretic framework. I will assume without discussion that *STR* is identified with the covert exhaustive operator *exh* proposed in Fox (2007), and that *ALT* is identified with the tree edit operations outlined in Fox and Katzir (2011). This means that the condition that any element $p \in ALT(S)$ needs to satisfy for it to become an actual implicature is that it needs to be 'innocently excludable' [as Fox (2007) defines the term; see below for illustrative examples]. This also means that alternatives are derived by substitution operations that replace focused nodes with subconstituents (for non-terminals) and with other lexical items (for terminals). My proposal about processing, however, will be compatible with different theories of *STR* and *ALT*; as noted above, the model I develop is concerned with the *inferences* that are generated, rather than the *mechanisms* that give rise to the inferences. This should make my proposal usable for scholars with other ideas about *STR* and *ALT* and their relation to the context of use.

Returning to (1), the basic/strengthened ambiguity follows from a systematic structural ambiguity: the sentence may or may not be parsed with *exh*. If *exh* is left off the parse, the sentence receives its basic meaning, and if *exh* is merged to the parse, the sentence receives its strengthened meaning. Following Fox (2007), the strengthened meaning of a sentence can often be paraphrased by adding *only* to the sentence (and focusing the relevant scalar item). Thus, *Jan ate only some of the cookies* and the strengthened meaning of (1) both convey (1-c). With both *exh* and *only*, $ALT(\exists) = \forall$ [by replacing *some* with *all* in (1)]. The question now is whether ∀ is innocently excludable. To test whether ∀ is innocently excludable, the mechanism negates it and examines whether the result of conjoining it with ∃ is consistent. The proposition ∃ ∧ ¬∀ is consistent, and hence the strengthened meaning ∃ ∧ ¬∀ is derived.

Innocent exclusion in this case was straightforward, but the mechanism is motivated by cases where non-trivial decisions need to be made about which alternatives to negate. Disjunctive sentences provide an illustrative example. Note that since *exh*

[1]Some proposals allow you to conjoin the basic meaning with unnegated alternatives (e.g., Chemla, 2009a; Bar-Lev and Fox, 2017). The differences between these theories will not concern us here (though see Note 24). What is important for current purposes is that strengthening occurs by conjoining a sentence with some other propositions derived from a restricted set of alternatives.

is general, it can apply to any sentence: $S \rightarrow exh(S, ALT(S))$[2]. The classic inclusive-exclusive ambiguity in disjunction, then, can be accounted for by the presence or absence of *exh*: without *exh*, the sentence receives the basic inclusive meaning in (2-a), and with *exh* the sentence receives its strengthened exclusive meaning in (2-c) by denying the alternative that Mary ate cake and ice-cream (2-b).

(2) Maria ate cake or ice-cream
 a. Basic meaning: $p \vee q$ (inclusive disjunction)
 b. Scalar implicature: $\neg(p \wedge q)$
 c. Strengthened meaning: $(p \vee q) \wedge \neg(p \wedge q)$ (i.e., the exclusive disjunction $p \oplus q$)

The set of alternatives for (2) is richer than the set of alternatives for (1). Here, as in (1), we have an alternative derived by lexical substitution: *or* is replaced by *and* to yield the conjunction $p \wedge q$. However, unlike (1), we have alternatives derived by replacing the root node by its subconstituents p and q[3]. Thus, $ALT(p \vee q) = \{p, q, p \wedge q\}$. The computation of innocent exclusion is more involved than with (1). The goal is to find the maximal subset of $ALT(p \vee q)$ that could be consistently negated with $p \vee q$. We can't negate the entire set, for that would contradict $p \vee q$. There are two maximal consistent exclusions: (i) $\{p, p \wedge q\}$, and (ii) $\{q, p \wedge q\}$. It would be arbitrary to select one of these maximal consistent exclusions over the other. For example, what would justify the negation of p over the negation of q? The only proposition that appears to be non-arbitrarily excludable is $p \wedge q$. A possibly useful motivation behind this idea is to think of (i) and (ii) as two different "votes" for which propositions to exclude. The alternative $p \wedge q$ is the only one that every vote agrees on, and for this reason it might be thought to be "innocently" excludable. Thus, $p \wedge q$ gets negated by *exh*, and the strengthened exclusive disjunction meaning $(p \vee q) \wedge (\neg(p \wedge q))$ is derived.

When the alternatives to disjunctive sentences are *not* closed under conjunction, innocent exclusion can assign a *conjunctive* strengthened meaning to disjunctive sentences[4]. Fox (2007) argues that this is the solution to the "paradox" of free-choice inference (Kamp, 1973). I will return to discussion of free-choice and its relation to innocent exclusion in later sections of the paper. I turn my attention now to relating this set of competence-theoretic ideas to performance models.

[2]It is known that *exh* has a restricted distribution (e.g., Singh, 2008a,b; Chierchia et al., 2012; Gajewski and Sharvit, 2012; Fox and Spector, 2018; Enguehard and Chemla, 2019). A more accurate characterization, then, is that *exh* can apply to any sentence in which it is licensed. All the examples we consider in this paper are ones in which *exh* is licensed.

[3]There are other possibilities here depending on what is assumed about the underlying parse. For example, if the *or* in the LF of (2) disjoins NPs instead of sentences, we would replace the noun phrase *cake or ice-cream* by each disjunct. The end result is the same in this case.

[4]In such cases, $ALT(p \vee q) = \{p, q\}$. Let S_0 be $p \vee q$ and let A_1 be $\{p, q\}$. The first application of *exh* on S_0 is vacuous because neither p nor q is innocently excludable: $exh(A_1, S_0)$ is equivalent to $p \vee q$. Let S_1 be the sentence $exh(A_1, S_0)$, and consider the exhaustification of S_1: $exh(ALT(S_1), S_1)$. The alternatives here are $\{exh(A_1, p), exh(A_1, q)\} = \{p \wedge \neg q, q \wedge \neg p\}$. Both are innocently excludable, and hence $exh(ALT(S_1), S_1)$ is equivalent to $p \wedge q$.

1.2. Processing Costs

At any given stage of the conversation, participants will have to decide whether to merge *exh* (and hence all of its arguments) to the parse of the uttered sentence[5]. To reduce clutter, I will simply write $exh(S)$ and omit mention of other arguments that *exh* takes, like $ALT(S)$ and c. The hearer thus faces a disambiguation task: they can either parse the sentence as S and add meaning $[[S]]$ to context c, or they can parse the sentence as $exh(S)$ and add meaning $[[exh(S)]]$ to c. It is plausible to assume that the choice has performance-theoretic consequences, and in particular that strengthened meanings ought to be costlier to process than corresponding basic meanings. To derive the strengthened meaning of sentence S, the processor needs to do all the work needed to compute S and its basic meaning $[[S]]$, and in addition it needs to create $ALT(S)$, determine which elements of $ALT(S)$ are innocently excludable, conjoin these innocently excludable propositions with $[[S]]$, and—under the identification of *STR* with *exh*—a more complex structure needs to be produced as well (for metrics, see e.g., Miller and Chomsky, 1963; Frazier, 1985, and many others). It would not be unnatural to expect this extra work to be realized in performance difficulties (see Chemla and Singh, 2014a for detailed discussion). To a significant extent, this expectation is borne out, at least with respect to cases like (1) and (2). For example, compared with their basic meanings, the strengthened meanings in (1) and (2) tend to be delayed in reading times in matrix positions (e.g., Bott and Noveck, 2004; Breheny et al., 2006) and in embedded positions (e.g., Chemla et al., 2016), they are late to develop (e.g., Noveck, 2001), they trigger later target looks in eye-tracking (e.g., Huang and Snedeker, 2009), and they are less frequently computed under time pressure (e.g., Bott and Noveck, 2004), under cognitive load (e.g., De Neys and Schaeken, 2007; Marty et al., 2013), and in embedded positions (e.g., Chemla, 2009b; Crnič et al., 2015).

Suppose that we take the above results to broadly indicate that the parser has a harder time with the form-meaning pair $< exh(S), [[exh(S)]] >$ than with the form-meaning pair $< S, [[S]] >$. Ideally this would follow from a general parsing theory. For example, we might consider the idea that a form-meaning pair $\lambda_1 =< f_1, m_1 >$ is easier to process than a form-meaning pair $\lambda_2 =< f_2, m_2 >$ if f_1 is contained in f_2 and the computation of m_1 is an intermediate step in the computation of m_2. The challenge would be to motivate the principle from general performance considerations, perhaps along the lines of the traditional "derivational theory of complexity" (see e.g., Fodor et al., 1974 for classic discussion). The core idea would be that processing costs are a monotonically increasing function of

[5]Some people have argued that sentences are always parsed with *exh* (e.g., Magri, 2009, 2011; Crnič et al., 2015). The observation that sentences aren't always strengthened is accounted for by appealing to contextual domain restriction in the alternatives. Everything we say here could be suitably translated into such a framework. For example, let A be $ALT(\exists) = \{\forall\}$, and let B be the result of contextual pruning of $ALT(\exists)$: $B = \emptyset$. Thus, instead of comparing $\exists$ and $exh(A, \exists)$ we would compare $exh(A, \exists)$ and $exh(B, \exists)$. Because our concern is only with the *meanings* of candidates, and not with their forms/computational histories, the proposal here could readily accommodate the assumption that *exh* is mandatory (along with competing ideas about strengthening). I will continue to assume here that *exh* is part of the inventory of logical operators and that its application is optional.

syntactic/semantic computational complexity: if the generation of λ_i involves a proper subset of the computations needed to generate λ_j, then (*ceteris paribus*) the cost of processing λ_i will be less than the cost of processing λ_j.

There are reasons to doubt that this monotonicity principle is on the right track. First, it appears committed to the assumption that there is a stage at which the parser has considered $< S, [[S]] >$ as the analysis of the sentence but not $< exh(S), [[exh(S)]] >$. Although natural, other views are also conceivable. For example, under a serial model of processing a single reading is entertained at any given point in processing; if it is found to be undesirable (for whatever reason) it may be replaced by a different reading generated by the grammar. In the case under consideration here, one would have to assume that *exh* appears late in the parser's structure-building. However, one could just as well begin by trying to parse with *exh* and revising only if necessary. This consideration is perhaps even stronger under the assumption that the human sentence processing mechanism uses a parallel processor. Suppose that the parser builds all (or at least many) of the form-meaning pairs that can be assigned to the sentence in a given context, and then decides (or asks the context to decide) which of these to select. Under such a model, the parser will already have produced both the strengthened and unstrengthened meanings, and it is not clear why the strengthened form-meaning pair should have any greater cost associated with it than the unstrengthened pair[6]. Under either view, we would be left with a stipulated "ordering" of computations in need of justification.

More importantly, there is empirical evidence against the monotonicity principle. First, return to the comparison with *only*. Like with *exh*, merging *only* to sentence S adds new syntactic and semantic computations. However, *only(S)* is not hard in the way that $exh(S)$ is. For example, parsing/interpretation of $exh(S)$ is slower than *only(S)* (e.g., Bott et al., 2012), memory demands inhibit $exh(S)$ but not *only(S)* (e.g., Marty and Chemla, 2013), and under certain conditions preschool children can compute *only(S)* even though they cannot compute $exh(S)$ (e.g., Barner et al., 2011). The sentences $exh(S)$ and *only(S)* involve very similar syntactic and semantic computations. Nevertheless, $exh(S)$ appears to be systematically harder than *only(S)*.

Taken together, these considerations suggest that costs arise precisely when a listener *chooses* $exh(S)$ over S during disambiguation. When processing *only some*, you cannot choose to understand the sentence as if *only* were not present. When processing (1), you have the option to understand the sentence with and without *exh*. The choice matters, and it appears that the disambiguation mechanism pays some kind of penalty for having chosen $< exh(S), [[exh(S)]] >$ over $< S, [[S]] >$. This might be taken as evidence for a restricted version of the monotonicity principle that becomes relevant only when the parser has to choose among competing analyses of the sentence. This would then leave us with the challenge of motivating the monotonicity assumption from general processing considerations. However, we will soon see that even this restricted version faces empirical challenges. In particular, the generalization we started with is incorrect: it is not *in general* true that $< exh(S), [[exh(S)]] >$ is harder than $< S, [[S]] >$. For some constructions, the opposite is true: $< exh(S), [[exh(S)]] >$ is sometimes *less costly* than $< S, [[S]] >$.

1.3. A Puzzle: Scalar Diversity in Processing

Assume that the basic meaning of numerals is an "at least" reading (3-a), and that the "exactly" reading follows from strengthening [(3-b) and (3-c); see Spector (2013) and references therein for relevant discussion of the basic and strengthened meanings of numerals].

(3) Numerals: Sandy ate three of the cookies
 a. Basic meaning: that Sandy ate at least three of the cookies
 b. Scalar implicature: that Sandy did not eat at least four of the cookies
 c. Strengthened meaning: that Sandy ate at least three of the cookies and did not eat at least four of the cookies, i.e., that Sandy ate exactly three of the cookies.

The pattern is thus like with (1) and (2): there is a basic meaning that gets strengthened by *exh*. However, the similarity does not carry over into processing: the strengthened meaning (3-c) is not costly relative to the basic meaning (3-a) (e.g., Huang and Snedeker, 2009; Marty et al., 2013). In fact, Marty et al. (2013) found that there were *more* exactly-readings of numerals under high memory load than under low memory load. This is the exact opposite of "some-but-not-all" type implicatures, which are reduced under high memory load. Thus, burdens on memory resources have the opposite effect for numerals and scalar items like *some*: strengthened meanings are *increased* with numerals and *decreased* with scalars.

Free-choice inferences are another puzzling case. A sentence like (4) has a so-called free-choice inference that Sandy is allowed to eat cake and is allowed to eat ice-cream—Sandy is free to choose (Kamp, 1973). The free-choice inference $\Diamond p \wedge \Diamond q$ does not follow from the logical form $\Diamond(p \vee q)$ if "$\vee$" is an inclusive disjunction and "$\Diamond$" is an existential quantifier over possible worlds. It has been argued—for example, on the basis of its sensitivity to monotonicity—that the free-choice inference is a scalar implicature (e.g., Kratzer and Shimoyama, 2002; Alonso-Ovalle, 2005). Various mechanisms have been proposed for deriving (4-c) as the strengthened meaning of (4) (e.g., Fox, 2007; Chemla, 2009a; Franke, 2011; Bar-Lev and Fox, 2017). I will not discuss these here[7]; what is important is that the free-choice

[6] Emmanuel Chemla (p.c.) notes that even under a parallel model it is conceivable that the parser could sometimes decide to stop at the smaller $< S, [[S]] >$, and this could account for the average cost difference. Like with the serial model, much depends on the "order" in which *exh* is applied. For example, one could design a parallel parser that creates parses top-down such that $exh(S)$ and S are always in the set of possibilities together, among other choice points.

[7] Note that $ALT((4)) = \{\Diamond p, \Diamond q, \Diamond(p \wedge q)\}$ is not closed under conjunction. Hence, we expect recursive exhaustification to yield the conjunctive free-choice scalar implicature $\Diamond p \wedge \Diamond q$ (the reader can use note 4 to work this out).

inference follows the pattern in (4), and hence is broadly similar to the patterns in (1), (2), and (3).

(4) Free-choice: Sandy is allowed to eat cake or ice-cream
 a. Basic meaning: $\Diamond(p \vee q)$
 b. Scalar implicature: $(\Diamond p \rightarrow \Diamond q) \wedge (\Diamond q \rightarrow \Diamond p)$
 c. Strengthened meaning: $\Diamond p \wedge \Diamond q$.

It turns out that free-choice inferences do not display the processing costs associated with (1) and (2). For example, they are processed faster than and are preferred to their basic meaning counterparts (e.g., Chemla and Bott, 2014), they are more robust under embedding than (Chemla, 2009b), and they are readily computed by children (Tieu et al., 2016). Furthermore, conjunctive strengthenings of disjunctive sentences more generally display these properties: preschool children (e.g., Singh et al., 2016b; Tieu et al., 2017) and adult speakers of Warlpiri (Bowler, 2014) appear to robustly compute conjunctive strengthenings of disjunction[8].

Let us use "free-choice" to refer to any conjunctive strengthening of disjunction. The challenge we face now is to explain why exhaustification in free-choice and in numerals has the opposite processing consequences than exhaustification in *some* and *or*. This is yet further evidence for a kind of scalar diversity (van Tiel et al., 2016), which takes seriously the observation that scalar implicatures for different constructions sometimes have different properties. Of interest to us here is that we now have evidence for a peculiar competence-performance mismatch:

(5) Competence-uniformity and performance–induced-diversity (CUPID):
 a. Competence-uniformity: The *competence* system treats the ambiguities in (1)–(4) in a uniform way, characterized as the optional application of a covert operator *exh* that computes innocent exclusion.
 b. Performance-induced-diversity: In some cases *exh* speeds up processing ((3), (4)), and in other cases it slows down processing (1), (2).

The challenge is to formulate auxiliary assumptions that relate the output of the competence system with measures of processing difficulty such that CUPID is predicted and things no longer seem peculiar. Clearly, any assumptions committed to scalar uniformity in processing will not work. This rules out the monotonicity assumption we were examining earlier under which $< exh(S), [[exh(S)]] >$ is generally harder to process than $< S, [[S]] >$. It also rules out principles such as the "strongest meaning hypothesis" [e.g., Chierchia et al., 2012, with roots in Dalrymple et al. (1998)] or "charity" (e.g., Meyer and Sauerland, 2009—see also Chemla and Spector, 2011). The goal of this paper is to meet this challenge.

[8]Podlesny (2015) argues that similar facts in American Sign Language follow the same pattern (though cf. Davidson, 2013). The pattern in question here is that disjunctive sentences can receive a free-choice (conjunctive) strengthened meaning when their alternatives are not closed under conjunction (see Fox, 2007; Chemla, 2009a; Franke, 2011; Singh et al., 2016b) and note 4.

1.4. Accounting for CUPID

Previous attempts at accounting for scalar diversity in processing have invariably made reference to language-internal computations and thus in some sense deny CUPID as a challenge to be solved. For example, some accounts have argued that strengthening has a cost when it requires a *lexical substitution* (as in "some but not all") but not when it requires only constituent substitutions (as in free-choice; e.g., Chemla and Bott, 2014; van Tiel and Schaeken, 2017). The guiding intuition, as I understand it, is that constituents are more readily accessible (they are already in the workspace), whereas lexical substitutions are more costly because the lexicon is presumably less accessible than material you have already created (you need to go out of the workspace to find a new lexical item). These considerations do not extend in any straightforward way to numerals, since their alternatives are derived neither by sub-constituents nor by lexical replacements (the set of numbers is infinite, and hence the alternatives must be referencing the successor function).

Numerals also seem to pose a challenge for the computation-specific proposal in Bar-Lev and Fox (2017). Specifically, they argue that free-choice and scalar implicatures like "some but not all" are derived by two different strengthening computations: roughly, the one for free-choice asserts the truth of alternatives and is context-independent and the mechanism for scalar implicatures negates alternatives and is context dependent. They argue that this distinction can be used to motivate a difference in processing costs. However, so far as I can tell, numerals are like scalar implicatures in the relevant competence-theoretic respects but they nevertheless pattern with free-choice in processing patterns (see also Note 24).

The model in Singh et al. (2016b) also made reference to language-internal computations but it readily accounts for numerals. Specifically, the model considers sets of form-meaning pairs the grammar assigns to the input sentence, and posits two constraints that interact to resolve the ambiguity: one pertaining to the candidate *meanings* and their relation to context, and the other pertaining to the candidate *forms* and their relative complexity. The syntactic assumptions assume the existence of a covert exhaustive operator that furthermore has a special pressure against it. I will discuss this model in greater detail in section 3.1, where I will modify it in various ways in the development of my proposal.

What the above accounts have in common is that they all relate processing costs in one way or another with the strengthening mechanism itself. Here I will pursue a different strategy. I will assume that CUPID teaches us that the costs of exhaustification are *unrelated* to the derivational history of the form/meaning of the sentence. Suppose that the language faculty delivers propositions (sets of worlds) to context-sensitive external systems of thought and action. By focusing our attention on the content produced by the language faculty—rather than on the mechanisms it uses to compute the given content—we might be in better position to develop closer connections between processing costs and arguably non-linguistic tasks like concept learning, theory selection, and communication viewed as a system of information exchange governed by social norms (see Grice, 1967; Fodor, 1983; Chomsky, 1995 among others for relevant discussion). At the same time, the focus on semantic output and context change could make our parsing assumptions relevant to a broader class of theories of the underlying competence system.

The focus on sentence meanings and their relation to contexts allows us to restate the disambiguation problem facing the listener as follows:

(6) Disambiguation as optimal context update: Suppose sentence S is uttered in context c, and suppose that the grammar $\mathcal{G}$ assigns k form-meaning pairs to S: $\mathcal{G}(S) = \{< f_1, m_1 >, \ldots, < f_k, m_k >\}$. These give rise to a candidate set of output contexts $\mathcal{C} = \{c_1, \ldots, c_k\}$, where $c_i = c + m_i$ (context c updated by m_i). The listener's task is to select the optimal element of $\mathcal{C}$ as the output context.

This context-update perspective has been found useful in studies of non-determinism in various domains, including parsing (e.g., Fodor, 1983; Crain and Steedman, 1985) and presupposition accommodation (see especially Beaver, 2001; von Fintel, 2008). I hope that it may shed insights into exhaustification decisions as well. Here, I will not say much about the (presumably decision-theoretic) optimality criterion used by the parser in solving (6). Instead, I will focus on the *costs* the parser faces when it chooses to update c with a particular m_i. There are two costs that I will consider: (i) the *a priori* complexity of m_i as a standalone object, here measured by semantic complexity (see section 2), and (ii) how well m_i resolves relevant uncertainties in c, and hence how much relevant uncertainty it leaves in c_i, where I identify relevant uncertainty with a function of the number of cells m_i eliminates from the question-under-discussion in c (see section 3). The sum of these costs, I argue, solves the challenge raised by CUPID.

2. SEMANTIC COMPLEXITY

I will begin by pursuing an idea, to my knowledge first suggested in the context of implicature computation by Bott et al. (2012), that the semantic complexity of different pieces of information might be relevant to how hard they are to process. To make this precise, we need an analytic framework that would make clear predictions about how to order different pieces of information for complexity. It turns out that there are branches of mathematical inquiry examining the semantic complexity of propositional and quantificational meanings. Furthermore, these analytical ideas have found useful application in concept learning, which in turn is arguably similar to theory selection and more generally to the choice of one element over some others. Of particular interest is the argument that the semantic complexity of a concept is a good predictor of how easy or hard it is for participants to acquire it (see especially Feldman, 2000 and subsequent work, such as summarized in Piantadosi et al., 2016). These results might thus provide antecedent motivation for the idea that certain pieces of information are intrinsically harder for humans to process than others, and this might be relevant to ordering the costs associated with exhaustification decisions.

2.1. Boolean Complexity and Processing Costs

Boolean functions like disjunction and conjunction map sets of truth-values (elements in $\{0,1\}^D$ for any number D) to a truth-value (an element in $\{0,1\}$). For example, if $D = 2$, there are four possible combinations of truth-values: $\{11, 10, 01, 00\}$. If $D = 3$, there are eight possible combinations: $\{111, 110, 101, 100, 011, 010, 001, 000\}$. More generally, there are 2^D possible combinations of D truth-values. Call this *Boolean D-space*. A Boolean function maps Boolean D-space into $\{0,1\}$. For example, inclusive disjunction maps any element to 1 so long as the element contains at least one 1[9].

A *Boolean Concept* is the characteristic set of the corresponding Boolean function. A concept is simply a way of carving a domain of interest into those instances that it is true of and those that it is not. For example, *dog* divides the universe into positive instances (things that are dogs) and negative instances (everything else). Similarly, Boolean concepts in D-space divide the 2^D possible truth-value assignments into those that are mapped to true and those that are mapped to false. For example, in Boolean 2-space the positive instances of inclusive disjunction are $\{11, 10, 01\}$. Similarly, exclusive disjunction picks out $\{10, 01\}$, and conjunction picks out $\{11\}$. These concepts, of course, can be thought of as propositions (sets of worlds). For example, the disjunctive concept $p \vee q$ is that set of worlds in which either just p is true, just q is true, or both p and q are true. We will go back-and-forth between concept talk and proposition talk.

We are interested in examining the extent to which these semantic notions have some intrinsic complexity. When we think of, say, the truth-table method for depicting Boolean functions, it is not immediately obvious why one table should be more or less complex than another. However, there is a perspective—which has been fruitfully applied to empirical facts concerning concept acquisition (Feldman, 2000)—that associates each Boolean concept with an intrinsic complexity measure. The method relates the complexity of a Boolean concept with the *smallest* Boolean formula that can express the concept using negation, inclusive disjunction, and conjunction as primitive (Feldman, 2000)[10].

[9]More generally, one is interested in functions that map $\{0,1\}^D$ to $\{0,1\}^{D'}$. We will not pursue this more general framework (see e.g., Savage, 1976).

[10]Of course, different primitives will give rise to different complexity measures. For example, exclusive disjunction requires at least four literals in a language with just $\wedge, \vee, \neg$ [see e.g., (9-d) and (9-e)]. Note that negation is not 'counted' in the measure—the motivation for this is that p and $\neg p$ divide logical space in the exact same way (Feldman, 2000). If exclusive disjunction were a primitive, $\oplus$ say, then you could get away with just two literals. Different complexity measures could also be considered. For example, the current measure does not count operators; some other measures would, such as ones that associate Boolean functions with complexity measures relating to the size or depth of circuits that compute them (see e.g., Sipser, 1997). For current purposes, I will assume that the concept learning literature (in particular Feldman, 2000) provides sufficient motivation for assuming that the set of primitives assumed here is telling, as is the assumed complexity measure. Note also that morphologically simplex operators in natural language appear to be restricted to just these primitives (Katzir and Singh, 2013). For relevant discussion, see also Piantadosi et al. (2016), Buccola et al. (2018), and note 16.

(7) Propositional formula: Consider a set of atomic propositional formulae as given. Then the set of propositional formulae is defined recursively as follows:
   a. Any atom p is a formula.
   b. If p is a formula, so is $\neg p$.
   c. If p and q are formulae, so is $(p \wedge q)$.
   d. If p and q are formulae, so is $(p \vee q)$.

We will sometimes omit parentheses when there is no risk of ambiguity.

(8) The Boolean Complexity of a concept C is the length n of the smallest formula f that expresses C: $n = min\{|f'| : [[f']] = C\}$.
   a. $|f'|$ is the number of *literals* in formula f'.
   b. A *literal* is any atomic formula p or its negation $\neg p$.

(9) Examples:
   a. $|(p \vee q)| = 2$
   b. $|(p \vee \neg q)| = 2$
   c. $|(p \vee q) \vee (p \wedge q)| = 4$
   d. $|(p \vee q) \wedge \neg(p \wedge q)| = 4$
   e. $|(p \wedge \neg q) \vee (\neg p \wedge q)| = 4$
   f. $|p \wedge q| = 2$.

Clearly, there are many formulae that can express a particular concept. For example, (9-a) and (9-c) both express an inclusive disjunction. However, (9-c) can be simplified to (9-a) without loss of meaning, and (9-a) is the shortest formula that can express inclusive disjunction in Boolean 2-space. There has been significant interest in finding mechanical methods for simplifying propositional formulae (e.g., Quine, 1952, 1955; McCluskey, 1956 and much other work). We will not discuss these here. For our purposes, what is important is that unlike the inclusive disjunction expressed in (9-c), the exclusive disjunction meanings expressed in (9-d) and (9-e) cannot be further compressed (Feldman, 2000). That is, there is no shorter Boolean formula capable of expressing an exclusive disjunction. In this sense, then, exclusive disjunctions are essentially more complex than inclusive disjunctions. They are also more complex than conjunctions [cf. (9-f)].

These complexity results align with empirical observations about the complexity of concept acquisition (again, see Feldman, 2000 and extensive references therein). Specifically, concepts whose membership is determined by an exclusive disjunction (e.g., "pink or square but not both") are harder to learn than concepts whose membership is determined by inclusive disjunction ("pink or square, possibly both") and they are also harder to learn than concepts whose membership is determined by conjunction ("pink and square"). This finding suggests that the human mind struggles with exclusive disjunctions in a way that it doesn't with inclusive disjunctions or conjunctions.

Consider now the exhaustification of an inclusive disjunction in the adult state. This leads to an exclusive disjunction interpretation, which we now have reason to think is inherently more complex than its inclusive disjunction counterpart. One way to make sense of the greater difficulty in processing $exh(p \vee q)$, then, is that it results in a more complex meaning than $p \vee q$. Specifically, it is plausible to assume that the parser incurs a penalty when it chooses to select a complex meaning even though a simpler one was available:

(10) Boolean Complexity and processing costs during disambiguation: Suppose that the grammar $\mathcal{G}$ assigns k analyses to sentence S: $\mathcal{G}(S) = \{\lambda_1, \ldots, \lambda_k\}$, where each λ_i is a form-meaning pair $< f_i, m_i >$. Let $B(m)$ be the Boolean Complexity of meaning m. Then the *cost of selecting* $\lambda_i \in \mathcal{G}(S)$, $C(\lambda_i)$, is proportional to the Boolean Complexity of meaning m_i: $C(\lambda_i) \propto B(m_i)$.

Note that the formulation in (10) only predicts processing costs that arise from disambiguation *decisions*. It would apply, then, to saying why $exh(p \vee q)$ is more costly to process than $p \vee q$ when the speaker utters a disjunctive sentence *p or q*, but it would not say anything about the relative cost of processing *only(p or q)* because no disambiguation is involved. Given a candidate set $\mathcal{G}(s)$, (10) partially orders this set by considering the Boolean Complexity of the meanings of its elements; this ordering, in turn, predicts relative processing costs when the hearer selects one or other element from $\mathcal{G}(s)$. However, (10) says nothing about how the cost of processing an element in $\mathcal{G}(s)$ would relate to the cost of processing a form-meaning pair outside of this set. Note also that the measure is context-invariant and that it does not reference the computational history of the elements of $\mathcal{G}(s)$. All that matters is what the different meanings in $\mathcal{G}(S)$ are.

The relative complexity of an exhaustified binary disjunction extends to Boolean k-space for any k. To simplify our discussion of the general case, first note that in the binary case $[[exh(p \vee q, ALT(p \vee q))]] = (p \vee q) \wedge \neg(p \wedge q) \iff (p \wedge \neg q) \vee (\neg p \wedge q) = [[exh(p, C) \vee exh(q, C)]]$, where $C = \{p, q\}$. More generally, where P_k is a k-ary disjunction $p_1 \vee p_2 \vee \ldots \vee p_k$ and $C = \{p_1, \ldots, p_k\}$, it is easily shown that $[[exh(P_k, ALT(P_k))]] = [[exh(p_1, C) \vee \ldots exh(p_k, C)]]$ (i.e., "only p_1" or "only p_2" or … "only p_k")[11]. This meaning can be expressed as the disjunction of k propositions, each of which is a conjunction of k literals in which one literal is positive and the rest are negative: $(p_1 \wedge \neg p_2 \wedge \ldots \wedge \neg p_k) \vee (\neg p_1 \wedge p_2 \wedge \neg p_3 \wedge \ldots \neg p_k) \vee \ldots (\neg p_1 \wedge \ldots \wedge \neg p_{k-1} \wedge p_k)$. Thus, exhaustification of P_k not only strengthens the meaning of

[11]It is sometimes argued that a large number of alternatives needs to be accessed during exhaustification, and that this could lead to computational costs (e.g., Mascarenhas, 2014; Spector, 2016). Note that the possibility of embedded exhaustification provides a significant reduction in the number of alternatives that need to be considered. Here we have one k-membered set, which gets used k times in exactly the same way each time. A global exhaustification using innocent exclusion could derive the same results by closing C under conjunction and ignoring closure under disjunction (see results in Spector, 2016). In fact, combinatorial explosion is at its worst when all we can do is blindly search through the entire space. This is not necessarily so with *ALT*, because there is sufficient structure within *ALT* that a sophisticated reasoner could exploit. For example, when finding maximal consistent exclusions (Fox, 2007), as soon as you decide that $p \wedge q$ is excludable, say, you can automatically conclude that any alternative r in which $p \wedge q$ is a subformula is also excludable (because r entails $p \wedge q$). Thus, algorithms for solving innocent exclusion might be able to avoid "perebor" (brute-force exhaustive search, no pun intended; cf. Trakhtenbrot, 1984). I will thus continue to assume that only the output of the language faculty is relevant to cost considerations. If it turns out that the number of alternatives is relevant, our cost formulation will have to change.

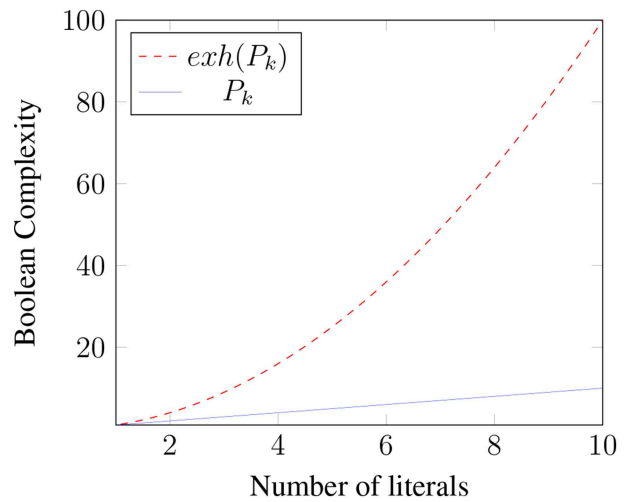


FIGURE 1 | Growth of Boolean complexity.

P_k, but it also creates a more *complex* meaning by converting a proposition with complexity k to one with complexity k^2. **Figure 1** illustrates how the Boolean Complexities of P_k and $exh(P_k)$ grow with k.

The Boolean Complexity perspective might thus provide a motivation for having *exh* in the first place. For note that *exh* allows speakers and hearers to convey relatively complex meanings by uttering relatively simple formulae. For example, *exh* allows speakers and hearers to use, say, a disjunction of 10 literals (hence complexity 10) to convey a message with 10 times that complexity: $B([[exh(P_{10})]]) = 100$. Of course, the application of *exh* also increases syntactic complexity (if we identify *STR* with *exh*), and the code for *exh* needs to be stored and executed. All of this will induce some cost. The tradeoff is presumably such that it is nevertheless an improvement on having to actually utter the more complex formula that would be required without *exh*.

Even if *exh* may have been "designed" in part to produce higher-complexity meanings from simpler ones, it does not always do so. For example, recall that under certain conditions *exh* can produce a *conjunctive* strengthening of a disjunctive sentence. Recall also that in such cases there appears to be no corresponding cost associated with *exh*. The Boolean Complexity analysis provides at least a partial answer to this: since conjunction and disjunction have the same Boolean Complexity, there is no expected cost under (10) when *exh* turns a disjunctive basic meaning into a conjunctive strengthened meaning[12].

Significant challenges remain. First, (10) does not speak to why conjunctive inferences should be *less costly* than their literal counterparts. Chemla and Bott (2014) found that—unlike scalar implicatures like "some but not all"—free-choice inferences are faster than their literal counterparts. They also found that—again unlike scalar implicatures like 'some but not all"—the rate at which free-choice inferences are selected does not drop under time constraints. As they put it (Chemla and Bott, 2014, p.392): "*not* deriving a free choice inference is a costly phenomenon." Furthermore, not only are conjunctive inferences less costly than their literal competitors, there appears to be a substantial *preference* to select the conjunctive reading when it is available (e.g., Chemla, 2009b; Bowler, 2014; Chemla and Bott, 2014; Meyer, 2015; Singh et al., 2016b; Bar-Lev and Fox, 2017; Tieu et al., 2017). In fact, even in concept learning, it is an old observation that conjunctive concepts are easier to acquire than disjunctive concepts. Thus, in both concept learning and in exhaustification, the order of difficulty appears to be the same:

(11) Cognitive difficulty of connectives: Conjunctions are easier than inclusive disjunctions which in turn are easier than exclusive disjunctions.

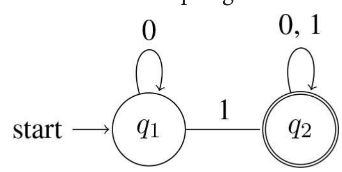
Boolean Complexity tells us why exclusive disjunctions are harder than inclusive disjunctions, but it does not tell us why inclusive disjunctions are harder than conjunctions. We will address this challenge in section 3. Before we do that, note that (10) is limited to propositional sentences. We need a general metric that could apply to quantified sentences as well. This would allow us to replace "Boolean Complexity" with a more general notion of "semantic complexity". We discuss this in the next section.

2.2. Semantic Automata

Consider sentences QAB, where Q is a quantifier, A its restrictor, and B its scope. Well-known constraints on natural language quantifier denotations allow us to view quantifiers as machines that determine acceptance/rejection based on two inputs only: those A that are B and those A that are not B (van Benthem, 1986). Call the first kind of input "1" and the latter "0." Given this perspective, quantifiers can be viewed as computational devices that accept certain strings over the alphabet $\{0, 1\}$. Call the set of strings accepted by the machine corresponding to quantifier Q the *language* accepted by Q, $\mathcal{L}(Q)$.

In the cases of interest to us, such as *some* and *all*, the quantifiers correspond to the simplest kinds of computing devices, namely finite-state-machines[13]. For example, a quantifier like *some* will accept any string as long there is at least one 1 in it (i.e., as long as there's at least one A that's a B). Here is a diagram of a machine that does this:

(12) Automaton accepting *some*:



[12] An interesting question is whether *exh* is always monotonic in semantic complexity. There is no logical necessity to this: a reviewer points out that $exh(p \oplus q, \{q\})$ means $p \wedge \neg q$, which is simpler than $p \oplus q$. The question of interest here is an empirical one: are there any cases of natural language sentences S such that $exh(S, ALT(S))$ has lower Boolean Complexity than S?

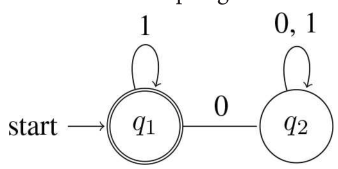
[13] Some quantifiers like *most* require push-down automata. There are close parallels between first-order definability and the Chomsky hierarchy. See van Benthem (1986).

In words, the machine starts in the start state q_1, and it processes the string one symbol at a time in left-to-right order. The arrows determine what the machine does upon processing a symbol. If it sees a 0 in state q_1, it remains in q_1 and moves on to the next symbol. If it sees a 1 in state q_1, it moves to state q_2 and moves on to the next symbol. Once in q_2, it remains there—neither a 0 nor a 1 can get it out of q_2. When all symbols in the string have been processed, the machine accepts the string if it is in an 'accept' state when the string ends; otherwise, it rejects the string[14]. In our diagram, q_2 is the 'accept' state, marked by double-circles.

Inspection of the machine in (12) at once reveals that it accepts strings like 1, 01, 000010101, 111, and that it rejects strings like 0, 000, and 00000. More generally, the language accepted by $\exists$ is $\mathcal{L}(\exists) = \{w : w \text{ contains at least one } 1\}$.

A quantifier like *all*, on the other hand, will reject a string as soon as it processes a single 0 (a single *A* that is not a *B*). That is, it accepts strings that contain only 1s: $\mathcal{L}(\forall) = \{w : w = 1^n \text{ for } n > 0\}$[15]. Here is a machine that accepts $\mathcal{L}(\forall)$ (note that in this machine, q_1 is both the start state and accept state):

(13) Automaton accepting *all*:



Given this formal apparatus, we can associate a quantifier *Q*'s semantic complexity with the size of the smallest machine that accepts $\mathcal{L}(Q)$:

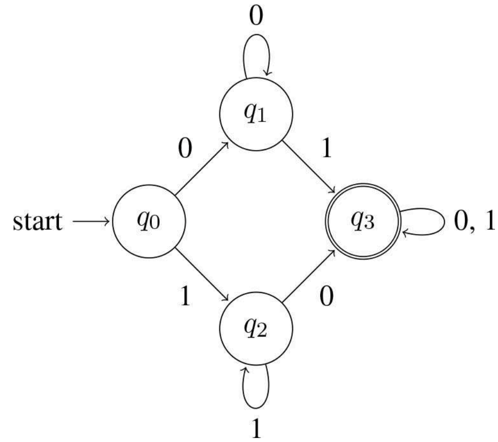
(14) Quantifier complexity:

a. The semantic complexity of a quantifier *Q* is the *minimum* size finite-state-machine that accepts $\mathcal{L}(Q)$.
b. The *size* of a machine is the number of states in the machine.

The machines in (12) and (13) are equally complex: they each have two states, and no smaller machines can be constructed that accept their respective languages. Note also that this definition of complexity is independent of the details of the syntactic expressions used to convey these meanings.

Now, recall that among the elements that $\mathcal{L}(\exists)$ accepts are strings like 11, 111, 1111, etc. These of course are the strings accepted by $\mathcal{L}(\forall)$. The semantic notion of entailment is realized here as a subset relation over bit strings: $\mathcal{L}(\forall) \subseteq \mathcal{L}(\exists)$. Application of *exh* breaks the entailment: $exh(\exists) = \exists^+ = \exists \wedge \neg\forall$, and $\mathcal{L}(\exists^+) = \{w : w \text{ contains at least one 0 and at least one 1}\}$. Here is a machine that accepts this language:

(15) Automaton accepting *some but not all*:



This machine is more complex than the ones in (12) and (13) (four states vs. two). Intuitively speaking, the additional complexity arises because determining membership in $\mathcal{L}(\exists^+)$ is a more demanding task. At any given point, a machine has to be ready to answer 'yes' or 'no.' Its memory is finite, but it does not know how long the input string is. Thus, the machine needs strategies for keeping track of relevant information without having to store the entire history of the string. The machine corresponding to $\exists$ in (12) needs to keep track of whether it has seen a 1 yet (if so, accept; otherwise, reject). The machine corresponding to $\forall$ in (13) needs to keep track of whether it has seen a 0 yet (if so, reject; otherwise, accept). The machine corresponding to $\exists^+$ in (15) needs to keep track of *both* of these pieces of information: it needs to keep track of whether it has seen a 1 yet and it needs to keep track of whether it has seen a 0 yet. The machine accepts the string only if the answer to both questions is "yes," but there are different paths to this state: one begins by having seen a 0 first, in which case the machine's strategy is to wait for a 1 and answer "yes" if and only if it encounters one, and the other begins by having seen a 1 first, in which case the machine's strategy is to wait for a 0 and answer "yes" if and only if it encounters one.

There is prior evidence that a quantifier's complexity has detectable psychological correlates. For example, recent evidence from implicit learning tasks suggests that concepts whose membership is determined by $\forall$ are preferred to those whose membership is determined by $\exists^+$ (Buccola et al., 2018). Like the relative ease of learning conjunctive concepts over exclusive disjunction concepts, considerations of semantic complexity

[14] More generally, a finite state machine is characterized by: (i) a finite set of states $\mathcal{Q}$; (ii) an alphabet Σ; (iii) a transition function $\delta : \mathcal{Q} \times \Sigma \rightarrow Q$ describing how the machine moves; (iv) a start state $q_1 \in \mathcal{Q}$; and (v) a set of accept states $\mathcal{F} \subseteq Q$. The machine in (12) has the following description: (i) $\mathcal{Q} = \{q_1, q_2\}$; (ii) $\Sigma = \{0, 1\}$, (iii) δ maps $(q_1, 0)$ to q_1, $(q_1, 1)$ to q_2, $(q_2, 0)$ to q_2, and $(q_2, 1)$ to q_2; (iv) q_1 is the start state, and (v) $\mathcal{F} = \{q_2\}$ is the (singleton) set of accept states. See any introductory text on formal language theory or the theory of computation for more detailed discussion of the properties of such machines (e.g., Sipser, 1997).

[15] A reviewer points out that (13) also accepts the empty string (the empty string is always accepted by machines for which the start state is an accept state). I omit mention of the empty string in the main text to avoid clutter and to simplify exposition. The reviewer notes that the machine here does not take existential import into account; without existential import, *all* would not entail *some*. The reviewer notes that a three-state machine would capture existential import. I believe we can sidestep the question of existential import because entailment is not needed for our purposes. As formulated in Fox (2007), *exh* negates not only stronger alternatives, but also those that are merely non-weaker. Either way, this will not affect our main point about the costs of strengthening *some* (though see note 16). I hope this makes it okay to ignore the empty string and its complications in the main text.

would appear to provide a natural account for this finding[16]. From a different direction, Szymanik and Thorne (2017) present evidence that the frequency of a quantifier's occurrence is to some extent predictable from its semantic complexity.

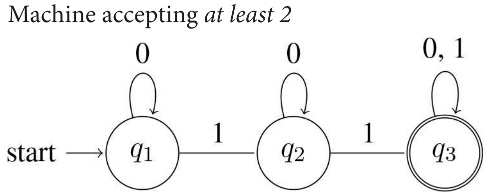
It is plausible, then, to think that quantifier complexity might also be a relevant factor in parsing costs. In particular, it might provide a rationale for why application of *exh* to $\exists$ tends to be costly: the meaning $\exists^+$ is inherently more complex than $\exists$ and is thus cognitively more demanding. Like with Boolean Complexity, the parser pays a penalty for choosing a complex meaning even though a simpler one was available.

(16) Quantifier complexity and processing costs during disambiguation: Let S_Q be a sentence containing quantifier Q, and suppose that the grammar $\mathcal{G}$ assigns k analyses to S_Q: $\mathcal{G}(S_Q) = \{\lambda_1, \ldots, \lambda_k\}$, where each λ_i is a form-meaning pair $< f_i, m_i >$. Let $Q(m)$ be the Quantifier Complexity of meaning m. Then the *cost of selecting* $\lambda_i \in \mathcal{G}(S)$, $C(\lambda_i)$, is proportional to the Quantifier Complexity of meaning m_i: $C(\lambda_i) \propto Q(m_i)$.

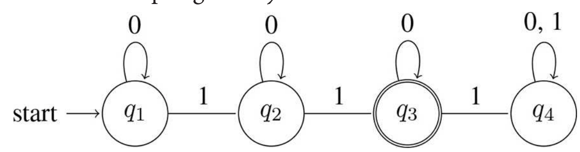
Given this definition, we will now simply use the term "semantic complexity" to refer to whichever of (16) or (10) applies, letting context choose.

Like with (10), the statement in (16) explains only some of the relevant facts. For example, consider numerals. A sentence like *Sandy ate two apples* on its basic meaning conveys that Sandy ate at least two apples. Its strengthened meaning is that Sandy ate exactly two apples. The strengthened meaning is not only stronger, but also more complex[17]:

(17) Machine accepting *at least 2*



(18) Machine accepting *exactly 2*



Despite this additional complexity, as we discussed earlier (section 1.3), the strengthened meanings of numerals are nevertheless easy to process and are often preferred to their unstrengthened counterparts. Furthermore, in concept learning the propositional results appear to carry over to their quantificational analogs. Specifically, it appears that when the data are consistent with $\forall$ and with $\exists$, learners tend to conclude that the underlying rule is universal rather than existential (Buccola et al., 2018). Clearly, semantic complexity cannot explain this[18].

In propositional and quantificational sentences, then, semantic complexity appears to provide at best a partial account of the relevant facts. In particular, it appears to explain cases where a disjunctive operator like $\vee$ or $\exists$ is strengthened by negating conjunctive alternatives like $\wedge$ or $\forall$, respectively. In such cases the result is a more complex meaning. In these cases, there is no CUPID: when the competence system applies *exh*, it creates a more complex syntactic object with a more complex meaning, and this complexity is realized in performance with a cost. It would make sense for there to be pressures to avoid this additional complexity if possible, and for there to be costs for selecting the more complex form-meaning pair against simpler alternatives. As noted, this pressure appears to be present in concept learning exercises as well: it is easier for participants to acquire a $\forall/\wedge$-concept than a $\exists^+/\vee^+$ concept.

However, semantic complexity does not speak to why $\forall/\wedge$-concepts are easier to learn/process than their $\exists/\vee$ variants. And semantic complexity does not explain CUPID: free-choice inferences and exactly-readings of numerals are less costly than their basic meaning counterparts even though they are not semantically simpler than them. Clearly, it can't be that semantically *stronger* meanings are less costly than their weaker basic meaning counterparts, given that $\exists^+/\vee^+$ are stronger than $\exists/\vee$ but are nevertheless harder to learn/process. The CUPID problem is still with us.

3. QUESTIONS, ANSWERS, CONTEXTS, AND PROCESSING COSTS

The above complexity measures provide an *a priori*, context-invariant ordering of meanings that the agent may apply before they have learned anything. As the agent accumulates information, and as the common grounds of their conversations become richer, these language-external domains will begin to exert a greater influence on parsing and interpretation strategies, and may in some cases counter the *a priori* orderings the organism starts with. I will argue that the solution to CUPID

[16] Buccola et al. (2018) conclude from their results that $\forall$, unlike $\exists^+$, is a plausible candidate for being a primitive in the language of thought. I will not enter into full discussion here, but it might be interesting to explore the connection between semantic complexity and logical primitives. If simplex lexical items are restricted to logical/conceptual primitives, then semantic complexity does not uniquely identify the primitives, given the existence of semantically simple but unlexicalized elements like *nand* (= $\neg\wedge$), *nall* (= $\neg\forall$), and others (e.g., Horn, 1972; Katzir and Singh, 2013). Furthermore, learnability arguments suggest that a logical primitives approach (as in Keenan and Stavi, 1986) can be dissociated from semantic complexity (see Katzir and Peled, 2018).

[17] More generally, machines accepting "at least n" require $n + 1$ states and those accepting "exactly n" require $n + 2$ states. Roni Katzir (p.c.) points out that the machine for "exactly n" could be simplified if we remove arrows leading to "sink states," i.e., non-accepting states like q_4 in (18) from which there is no escape. If we were to do this, the machines in (17) and (18) would have the same complexity. The empirical problem at hand would remain even if we adopted this way of counting: we need to account for the observation that the exactly-reading of numerals appears to not only be free (relative to its basic meaning counterpart), but in fact less costly than it (Marty et al., 2013). I will thus continue to assume that sink states are included in the complexity measure, and hope that we could restate things if it turns out that removing them would be preferred.

[18] Nor does the alleged primitiveness of $\forall$, assuming that $\exists$ is also primitive (cf. Note 16).

involves considerations of how candidate meanings interact with the context of use. On the classic Stalnakerian picture, sentences are uttered and understood in context, and sentences update the input context in rule-governed ways to create a new output context relative to which the next utterance will be interpreted. Thus, contexts and sentences have a dynamic interplay that we will momentarily exploit to help us overcome the limitations of semantic complexity alone.

Specifically, I will argue (building on Singh et al., 2016b) that the extent to which a given meaning resolves the question under discussion (QUD) is a predictor of the costs of accepting it into the common ground. The better the answer, the lower the cost. Here, "goodness" is a function of how close to a complete answer the meaning provides, i.e., how close it comes to locating the one true cell in the partition induced by the QUD. I motivate this idea briefly in section 3.1, and show in section 3.2 that the parsing mechanism proposed in Singh et al. (2016b) provides the pieces needed to overcome the problem posed by numerals and free-choice inferences. That system included two interacting constraints that were evaluated by an Optimality-Theoretic system: (i) a constraint that penalizes incomplete answers (considered as semantic objects), (ii) a constraint that penalizes syntactic complexity (occurrences of *exh*). In section 2, I proposed a way to replace (ii) with a measure of semantic complexity, and in section 3.2, I show how to incorporate this amendment into the system in Singh et al. (2016b).

In sections 3.3 and 3.4, I further modify Singh et al.'s (2016b) proposal by changing the way processing costs relate to answers. Specifically, Singh et al. (2016b) suggested that complete answers have no cost but partial answers do, and that partial answers are equally costly. In section 3.3, I will motivate the idea that partial answers can be ordered for quality by how far they are from complete. I also review and reject some simple options for formalizing this distance, and in section 3.4 I provide a domain-general way to measure distance using the information-theoretic concept of entropy (Shannon, 1948). Entropy has a well-known compression interpretation (number of bits needed to eliminate the uncertainty), thus making it plausible that both semantic complexity and entropy have a compression-related cost. I will suggest that this lends flexibility in formulating functions that combine these costs. For example, it allows us to abandon the OT evaluation system and instead use simple arithmetic. Here is my proposal:

(19) Processing costs during disambiguation: Let S be a sentence uttered in context c. Suppose that grammar $\mathcal{G}$ assigns k analyses to S: $\mathcal{G}(S) = \{\lambda_1, \ldots, \lambda_k\}$, where each λ_i is a form-meaning pair $< f_i, m_i >$. Let $\mathcal{S}(m_i)$ be the semantic complexity of m_i, let c_i be the result of updating context c with m_i, $c + m_i$, and let $\mathcal{H}(c_i)$ be the entropy in context c_i. Then the *cost of selecting* $\lambda_i \in \mathcal{G}(S)$ in context c, $C(\lambda_i, c)$, is: $C(\lambda_i, c) = \mathcal{S}(m_i) + \mathcal{H}(c_i)$.

We will now build our way to the cost function in (19), highlighting various choice points as we go. We begin with the importance of questions and answers and more generally with the way normative demands on speech might play a role in processing costs.

3.1. Norms of Good Conversational Behavior and Processing Costs

It is commonly assumed that there are *normative* demands on a speaker, such as the demand that they be truthful, informative, relevant, assert things they have evidence to support, use sentences whose presuppositions are satisfied (or easily accommodated), among other constraints on their behavior (e.g., Grice, 1967; Stalnaker, 1978; Williamson, 1996, and much other work). Listeners pay attention to whether these demands are satisfied. There are consequences when it is detected that a speaker misbehaved according to these norms. There is surprise, embarrassment, hostility, and trust and credibility are broken. These considerations suggest that the maxims should be viewed as rules of decent cooperative behavior, which in particular apply even when it is in the speaker's interest to violate them. A speaker may decide, for instance, to speak a falsehood or omit relevant damning information, but even if this maximizes their utility in some sense this would not justify their action. They are held to the maxims independent of the utility of their doing so. All else being equal, then, we assume that a speaker is more likely to be obeying the norms than violating them.

(20) Assumption about language use: Unless we have reason to think otherwise, assume that a speaker is obeying conversational maxims.

If (20) is a true assumption about conversation, we would expect it to be relevant to disambiguation. In particular, suppose that λ_1 and λ_2 are competing form-meaning pairs, and that λ_1 violates a norm of language use and λ_2 does not. We would expect (20) to generate a pressure in favor of λ_2. It is of course hard to tell whether someone is speaking truthfully, or has evidence to support what they assert. But it is easy to tell whether a speaker is being *relevant*[19]. Specifically, suppose that the ideal speaker is assumed to be optimally relevant, by which we mean that they immediately (when it's their turn to speak) settle the Question Under Discussion (QUD). Assume further that QUDs can be modeled as partitions of the common ground (e.g., Groenendijk and Stokhof, 1984; Lewis, 1988, among others). For example, $PART(c) = \{pq, pq', p'q, p'q'\}$ is a partition that divides c into four sets of worlds (cells of the partition): those where p and q are both true (pq), those where p is true and q is false (pq'), those where p is false and q is true ($p'q$), and those where both p and q are false ($p'q'$). An *answer* is a union of cells, and a *complete answer* is a particular cell.

What we want in a context is a complete answer. If I ask you who was at the scene of the crime, and you know the answer ('the whole truth'), you are required to tell me. Given any proposition r asserted by the speaker, we can readily examine whether r—together with the information in the common ground—identifies a cell. That is, we can readily answer the question: $\exists u \in PART(c) : u = r \cap c$? If the answer is positive, the listener will be satisfied that the question has been resolved. Otherwise, the

[19]It is also easy to tell whether the uttered sentence's presupposition is satisfied or is otherwise innocuous (just compare the presupposition with the information in the common ground). When it is not, there are detectable and immediate costs for accommodation (e.g., Singh et al., 2016a).

speech act will have left undesired relevant uncertainty. This goes against our expectation that the speaker would fulfill their obligations, at least if they don't flag that they are unable to do so.

Thus, consider the following principle proposed in Singh et al. (2016b)[20].

(21) Complete Answer Preference: If there is an analysis $\lambda_i = < f_i, m_i >$ of sentence S such that m_i completely answers the QUD in c, then—all else being equal and assuming no other candidate completely answers the QUD—λ_i will be preferred.

Suppose, then, that the parsing mechanism encodes an expectation that the speaker is obeying all relevant maxims. The parser will therefore expect to find among the form-meaning pairs provided by the grammar one that will completely answer the QUD (among other demands on good conversational behavior). If it finds one, then it will select it and no cost is induced. They have simply applied their grammatical principles to analyze the sentence and their normative expectations have been satisfied. However, something goes wrong if the QUD is not completely answered. The listener will be surprised, and other considerations might enter into disambiguation decisions and therefore also into the consequences of these decisions.

3.2. The Parsing Proposal in Singh et al. (2016b) With Semantic Complexity in Place of Syntactic Complexity

Singh et al. (2016b) suggested an Optimality-Theoretic processing mechanism that incorporated a preference for a complete answer and a pressure against syntactic complexity. Specifically, the system posited (i) a high-ranked constraint *INC that penalizes form-meaning pairs that fail to provide a complete answer to the QUD, and (ii) a low-ranked constraint *exh that penalizes a form-meaning pair for each occurrence of *exh* in the parse. In that system, when no form-meaning pair provides a complete answer to the QUD, considerations of syntactic complexity (approximated by number of occurrences of *exh*) adjudicate between the remaining candidates. By ranking *INC above *exh, the system assumes that a sentence's ability to resolve relevant contextual uncertainty is worth any syntactic cost that might be incurred by adding *exh*. Furthermore, by positing *exh, the system identified the number of occurrences of *exh* as a proxy for the sentence's complexity, and hence used the *form* of the sentence as its complexity measure.

In this paper I am pursuing the idea that the parser is only sensitive to the *meanings* of candidates. Thus, when no form-meaning pair provides a complete answer, the amendment needed in Singh et al. (2016b) would be to posit that *semantic* complexity determines the parser's choice. This could be implemented by replacing *exh with *SC (for "semantic complexity"), and by assigning a candidate form-meaning pair a number of violations equal to its semantic complexity. Here we show that this amendment captures all the facts that Singh et al.'s (2016b) proposal was designed to account for, and that the constraint *INC accounts for CUPID under the assumption that it is higher-ranked than *SC.

Consider again the question faced by a listener about whether or not to exhaustify the input sentence. Suppose that a disjunctive sentence like $p \vee q$ is uttered in response to a (possibly implicit) QUD like *which of p and q is true*? That is, suppose it is uttered in a context in which the partition is $PART(c) = \{pq, pq', p'q, p'q'\}$[21]. Of course, a disjunctive answer $p \vee q$ only gives a partial answer, eliminating just the cell $p'q'$. A better answer is made available by *exh*: in the adult state $exh(p \vee q)$ would also eliminate the cell pq. This is better—it generates fewer ignorance inferences than the parse without *exh* (Fox, 2007)[22]. However, it is still an undesirable and unexpected state of affairs because it continues to leave us with relevant uncertainties. In fact, as noted in Singh et al. (2016b), we appear to have prosodic contrasts between complete and partial answers, but not between better and worse partial answers. This observation indicates that what matters for answerhood—at least so far as prosody is telling—is whether the sentence provides a path to a complete answer. In the adult state with plain disjunctive sentences, the parser has no analysis available to it that provides it with a complete answer. In such a case, *SC will get a chance to decide the optimal analysis. Here, the *a priori* ordering between the simpler inclusive disjunction and the more complex exclusive disjunction (cf. section 2.1) would pressure against the exclusive disjunction. Assuming that less optimal candidates are costlier than optimal candidates, we predict the observed cost for the exclusive disjunction reading of *A or B*.

(22) Strengthening inclusive disjunction to exclusive disjunction:

		A or B	*INC	*SC
a.	☞	<A or B, $A \vee B$ >	*	**
b.		<exh(A or B), $A \oplus B$ >	*	****

Things are different when disjunctive sentences have alternatives that are not closed under conjunction. In such cases, *exh* can turn the disjunction into a conjunction (Fox, 2007; Singh et al., 2016b; see also Chemla, 2009a; Franke, 2011; Bar-Lev and Fox, 2017 and note 4). Assume the treatment in Fox (2007) and Singh et al. (2016b) under which recursive application of *exh* turns $p \vee q$ into $p \wedge q$: $[[exh^2(p \vee q)]] = p \wedge q$. On the face of it one might have expected this computation to be hard, since there are multiple applications of *exh* and multiple sets of alternatives that get generated. However, recall that we are assuming that these

[20]See also Katzir and Singh, 2015 for a related but somewhat different notion of the "goodness" of answers, together with suggestions about the goodness of questions as well.

[21]There might be more propositional variables under consideration, but this doesn't affect anything we have to say here.

[22]Fox (2007) notes that the pure Maxim of Quantity leads only to ignorance inferences about all relevant propositions whose truth-values are not settled by the speaker's utterance. This in turn follows from considerations of relevance (the so-called 'symmetry problem'; cf. von Fintel and Heim, 1999). In Fox's (2007) system, exhaustivity is a mechanism that helps conversational participants take sentences that are at best partial answers and convert them into better partial answers or into complete answers where possible (see especially Fox, 2018 for extensive discussion with consequences for the semantics and pragmatics of questions more generally).

computations do not contribute to costs. Instead, it is the *output* of these computations (*SC), and its affect on the context (*INC), that are relevant to processing costs. In this case, the parser finds the conjunctive meaning and considers it desirable because it provides a complete answer and no cost is therefore expected.

(23) Strengthening inclusive disjunction to conjunction:

		A or B	*INC	*SC
a.		<A or B, $A \vee B$ >	*	**
b.	☞	<exh(exh(A or B)), $A \wedge B$ >		**

More generally, if it is reasonable to assume that a disjunction $P_k = p_1 \vee p_2 \vee \ldots p_k$ will typically be used in a context in which the participants are interested in knowing, for each of the disjuncts p_i, whether p_i is true, then we have an explanation for the contrast between conjunctive strengthenings and exclusive strengthenings and their relative ordering with inclusive disjunction [cf. the generalization in (11) in section 2.1]: conjunctive readings satisfy the high-ranked **INC* whereas neither inclusive nor exclusive disjunctions do, and inclusive disjunctions have fewer **SC* violations than exclusive disjunctions.

The result extends to quantificational sentences *DAB*, where *D* is a quantificational determiner, *A* its restrictor, and *B* its nuclear scope. Suppose such sentences are typically used in answers to the question *How many A B?* If there are *k* individuals in the domain, then this induces a partition with $k + 1$ cells ("none," "exactly 1," "exactly 2," ..., "exactly k"). When *D* is a logical existential quantifier as in *some A B*, the basic meaning $\exists$ only eliminates the "none" cell. This partiality is expected, given that existential quantifiers are essentially disjunctive: "exactly 1 or exactly 2 or ... or exactly *k*." Exhaustification can produce a slightly better answer by eliminating the "exactly k" cell, but it still typically leaves you without the expected and desired complete answer because you are still left wondering which of exactly 1 or exactly 2 or ... or exactly $k - 1$ is true. Thus, both $\exists$ and $\exists^+$ violate **INC*. However, because $\exists$ is semantic simpler than $\exists^+$ (2 vs. 4; cf. section 2.2), **SC* decides in favor of $\exists$ and $\exists^+$ is therefore predicted to be costly.

If the question were one that induced a different partition, say $\{\exists \wedge \neg\forall, \forall, \neg\exists\}$, then the costs for *exh*($\exists$) could disappear because it would now satisfy the high-ranked **INC* and $\exists$ still would not (see Breheny et al., 2013 for evidence in this direction). This is a general feature of the proposal: the costs for processing any sentence *S* will depend on what the QUD is. Sometimes *exh* can help you turn *S* into a complete answer, in which case no cost is expected, but other times *exh* will only create more complex meanings without also creating a complete answer, in which case costs are expected[23].

When *D* is a numeral, *exh* will typically produce a complete answer to a *how-many* question. Suppose that there are *k* individuals in the domain, and that the speaker produces *nAB* where $n < k$. On its basic meaning, *nAB* is again only a partial answer, eliminating all cells "exactly r" where $r < n$. Again, this is expected given that the basic meaning is essentially disjunctive: "either exactly *n* or exactly $s(n)$ or ... or exactly $s^j(n)$" (where $s^j(n) = k$ and is the result of $j = k - n$ applications of the successor function to *n*). But with numerals, unlike with logical *some*, *exh* can produce a complete answer by also eliminating cells "exactly r" where $n < r \leq k$ (because, following Horn, 1972, the alternatives for *n A B* include not just *k A B*, but also *r A B* for $n < r \leq k$)[24]. For example, consider the case where *n* = 2. Refer to the basic "at-least" reading with $[\geq 2]$, and to the strengthened "exactly" reading with $[= 2]$. Then the OT constraint evaluation system selects $[= 2]$ as optimal because it satisfies **INC*, even though the "exactly" reading incurs more violations of the lower-ranked **SC* (cf. Note 17 in section 2.2):

(24) Strengthening numerals from an "at least" to an "exactly" reading:

		2*AB*	*INC	*SC
a.		$< 2AB, [\geq 2] >$	*	***
b.	☞	$< exh(2AB), [= 2] >$		****

The system in Singh et al. (2016b) thus accounts for CUPID by appealing to the importance of complete answers in an overall theory of processing costs. The complete answer perspective may also speak to some of the questions that remain unanswered in concept learning. Recall that conjunctive concepts are easier to learn than inclusive disjunction concepts, and that universal quantification is easier to learn than existential quantification. We now have a rationale for this: if you learn that some element satisfies a conjunctive concept (say *red and triangle*), you learn right away that it is red and that it is a triangle. Disjunctive concepts—whether inclusive or exclusive—leave this question open. Similarly, learning that *All wugs are red* tells you that as soon as you encounter a wug, you can infer something about its color. Learning only that some wugs are red, or that only some wugs are red, does not confer you with this inferential ability. Presumably, as with conversation, it is better to have relevant uncertainties resolved than to leave them unresolved. Recall

[23] A reviewer raises the question of how we can identify the QUD of an utterance. For example, consider a context in which the goalkeeper Sue must not let in more than 2 goals to keep her position as starting keeper. A asks: *Did Sandy keep her position?* B responds: *No, she let in three goals.* In a sense, the strengthened meaning of B's response gives strictly more information than is required to answer A's question, whereas the basic meaning itself gives exactly the right amount. The reviewer wonders whether the QUD for *she let in three goals* might nevertheless be a *how many* question. There is certainly room for flexibility of QUDs, and numerals might strongly be associated with *how many* questions. At the same time, we have not said anything about how to incorporate an overly strong answer in our measure of "distance from a complete answer." I leave this as a challenge for now.

[24] Bar-Lev and Fox (2017) propose that "innocent inclusion"—a new method for computing free-choice—is obligatory and hence cost-free while "innocent exclusion" (Fox, 2007)—used for more standard scalar implicatures (like $\exists^+$)—has a cost due to context-sensitive optionality. The case of numerals suggests that complete answerhood is the more fundamental notion. Of course, this does not speak at all to the motivation for introducing innocent inclusion in the first place (the need for a global mechanism to compute universal free choice—Chemla, 2009b).

that these considerations cannot be reduced to considerations of semantic strength: for example, conjunction and exclusive disjunction are both stronger than inclusive disjunction, but only conjunction is easier to process.

3.3. Complete vs. Partial Answers

We have been assuming with Singh et al. (2016b) that the parser cares only about whether a given form-meaning pair provides a complete answer to the QUD. As we noted earlier, this assumption is motivated in part by the observation that our pronunciation patterns distinguish between complete and partial answers but not between different kinds of partial answer. An additional motivation comes from considerations of our obligations in general. If I ask my son to help me carry a stack of books from one room to the other, and the request is reasonable, I expect him to help me move all of them. I would be surprised and disappointed with anything less.

But what if he helped me move half of them and then went back to his video games? Is that not better than opting out entirely? The system in Singh et al. (2016b) treats all sub-optimal answers on a par. For example, in (22) both the inclusive disjunction and exclusive disjunction receive a single penalty for violating **INC*, even though the exclusive disjunction is a better answer (it rules out two cells instead of only one). Even if prosody is blind to this distinction, it is not obvious that the parsing mechanism should be. Some partial answers are closer to complete than others, and it is conceptually natural to think that the parser might care about how close different possibilities get to the end goal. To facilitate comparison with Singh et al.'s (2016b) binary choice (*complete or not?*), it would be useful to formulate a measure that allowed partial answers to be compared for how far they are from complete. Here we aim to find such a measure, and to examine its usefulness in accounting for the facts under discussion. Here I review some fairly simple measures, but I will reject them in favor of the information-theoretic entropy measure proposed in section 3.4. Readers may skip straight to the proposal there, but I provide details here because it might be instructive to see why arguably simpler proposals don't work.

One natural amendment of Singh et al. (2016b) that could accommodate the ordering assumption would be to count the number of remaining cells in the partition and to use that as the number of **INC* violations (1 being the minimum value associated with the complete answer). Call our new constraint **INC-G* (where *G* is for "graded"). Under this view, conjunctions would still be optimal when compared with inclusive disjunctions: they identify a unique cell, whereas disjunctions leave three cells to choose from.

(25) Strengthening inclusive disjunction to conjunction:

A or B	*INC-G	*SC
a. $<$A or B, $A \vee B>$	$***$	$**$
b.☞$<$ *exh*(*exh*(A or B)), $A \wedge B>$	$*$	$**$

Unfortunately, the move from **INC* to **INC-G* quickly runs into trouble. For example, exclusive disjunctions come out as optimal in competition with inclusive disjunctions because they only leave behind two cells:

(26) Strengthening inclusive disjunction to exclusive disjunction:

A or B	*INC-G	*SC
a. $<$A or B, $A \vee B>$	$***$	$**$
b.☞$<$exh(A or B), $A \oplus B>$	$**$	$****$

This is the wrong result. We could correct for this by actually reordering the constraints such that **SC* outranks **INC*. This would work for (26) and for (25), but it would not work for numerals. For example, if the sentence *2 AB* is offered in response to the question *how many (of these 4) As are B?*, the evaluation component would select the basic "at-least" reading as optimal:

(27) Strengthening numerals from an "at least" to an "exactly" reading:

2*AB*	*SC	*INC-G
a. ☞ $<2AB, \geq 2>$	$***$	$***$
b. $<exh(2AB), =2>$	$****$	$*$

These considerations could of course be taken as an argument that the parser does not after all distinguish between different kinds of partial answer, and thus that the parsing mechanism incorporates **INC* instead of **INC-G* and orders **INC* over **SC*. The challenge for this view would be to provide a rationale for why the constraints should be ordered in this way.

In the rest of this paper I will continue to take a different path so that we have a concrete viable alternative that allows room for orderings of partial answers. As a starting point, suppose that the problem is not with **INC-G* but with the OT evaluation system. Specifically, assume that costs are equated with the total number of constraint violations. Different cost functions are imaginable, but let us take summation as a simple starting point. Under this view, it turns out the above facts can all be captured. For example, in the case of binary connectives, conjunctions are less costly than inclusive disjunctions (three vs. five) which in turn are less costly than exclusive disjunctions (six). Similar results hold for quantified sentences. Suppose that there are k individuals in the domain. Then $\exists$ costs $k+2$ and $\exists^+$ costs $k+3$: $\exists$ incurs two violations of **SC* and k violations of **INC-G* (it only eliminates the cell in which no individuals that satisfy the restrictor satisfy the scope, leaving behind k cells), and $\exists^+$ incurs four violations of **SC* and $k-1$ violations of **INC-G* (it also eliminates the cell in which all individuals that satisfy the restrictor satisfy the scope). Finally, numerals nAB (where $n < k$) are also accounted for: the "at-least" reading has $n+1$ violations of **SC* and $(k-n)+1$ violations of **INC-G*, and hence $k+2$ violations in total, whereas the "exactly" reading has $n+2$ violations of **SC* and one of **INC-G*, for $n+3$ violations in total. For all values of n and k such that $n < k$, the "exactly" reading is no more costly than the "at least" reading, and for all but the case $k = n+1$ the "exactly" reading is less costly.

Unfortunately, this perspective leads to some counter-intuitive predictions. Consider the case of a general k-ary disjunction P_k, and consider the costs associated with the basic meaning of the sentence, as well as with $exh(P_k)$ (leading to the "only one" reading) and with $exh^2(P_k)$ (leading to the conjunctive reading when the alternatives are not closed under conjunction; see note 4). Recall from section 2.1 that the semantic complexity of P_k is k (the smallest formula representing this meaning is $p_1 \vee p_2 \vee \ldots \vee p_k$), which is also the semantic complexity of $exh^2(P_k)$ because this gives the incompressible $p_1 \wedge p_2 \wedge \ldots p_k$. Recall also that $exh(P_k)$ is more complex: its meaning is given by k disjuncts each of which contains k conjuncts that assert that one of the p_i is true and all other $k-1$ p_j are false. Thus, $exh(P_k)$ has semantic complexity k^2. We also need to say something general about how these meanings affect the QUD. With k literals, there are 2^k cells of the partition. Conjunctions completely answer the QUD, and hence leave behind a single cell in which each literal p_i in P_k is true. Inclusive disjunctions P_k eliminate only the cell in which all literals p_i in P_k are false, and hence they leave behind $2^k - 1$ cells. Finally, $exh(P_k)$ leaves behind k cells in each of which only one of the literals p_i in P_k is true. We summarize these costs in (28):

(28) Costs of update (to be revised):

Formula	$*SC$	$*ING\text{-}G$
P_k	k	$2^k - 1$
$exh(P_k)$	k^2	k
$exh^2(P_k)$	k	1

Continue to assume that costs are simply added together. The cost of the conjunctive reading grows linearly with k (it is the sum $k + 1$), and thus still comes out less costly than the disjunctions because their costs grow more rapidly: $exh(P_k)$ grows as a polynomial $k^2 + k$, and P_k grows exponentially $k + 2^k - 1$. The competition between the two disjunctions thus boils down to how quickly k^2 grows vs. $2^k - 1$. It turns out that exclusive disjunctions are predicted to be slightly more costly than inclusive disjunctions for $2 \leq k \leq 4$ (in this range $2^k - 1 < k^2$), after which point the costs of inclusive disjunctions start to increasingly dwarf the costs of exclusive disjunctions (here $2^k - 1 \gg k^2$). See **Figure 2** for an illustration.

It would be surprising, hence interesting, if this prediction were true. But it seems rather unlikely. A more natural result would be one under which inclusive disjunctions are truly sandwiched between exclusive disjunctions and conjunctions for all values of k. Certainly, this is what all the evidence would suggest (Feldman, 2000). The problem, clearly, is the exponential cost associated with inclusive disjunctions because of the poor job they do at answering questions. They eliminate only one among an exponential space of cells, and they therefore leave behind an exponentially large amount of relevant uncertainty.

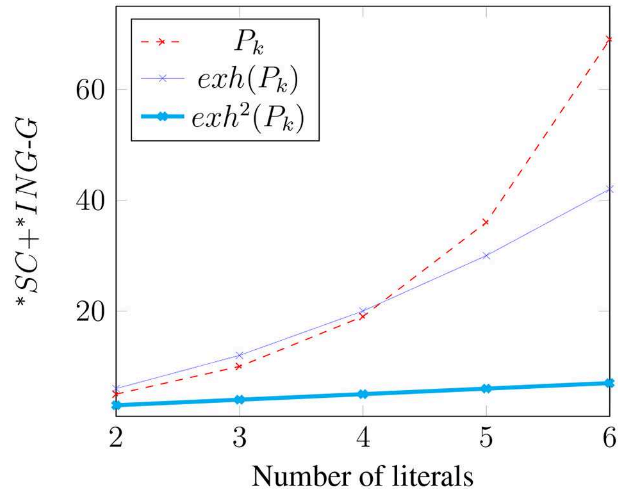


FIGURE 2 | Cost of update (to be revised).

3.4. Entropy, Questions, and Answers

There is a natural perspective that tames the costs associated with exponential relevant uncertainty (van Rooij, 2004, building on Bar-Hillel and Carnap, 1952 among other work). Suppose that we identify relevant uncertainty with the *entropy* of a partition, which measures the amount of information a receiver would expect to receive from observing an outcome of this partition. Your relevant uncertainty is eliminated when you observe a given outcome, and the entropy of the partition therefore provides a natural measure of the amount of relevant uncertainty you started with. Clearly, the greater the number of alternatives we are considering, the more relevant uncertainty there is and hence the more informative any particular outcome would be. We thus want a measure of relevant uncertainty that is monotonically increasing in the number of cells. Simply counting the number of cells provides such a measure but as we saw it runs into trouble. Note also that the count measure makes no use of probabilities. For example, there is a sense in which a less likely cell is more informative than a more likely one. There is also a sense in which we are most uncertain if all cells are equally likely.

To account for these and other desiderata, Shannon (1948) argued that the information associated with any given cell q_i in partition $\mathcal{Q}$ should be identified with $log(1/P(q_i))$, where $P(q_i)$ is the probability that q_i is the answer to the question (the message that we receive). From this, the relevant uncertainty of the partition is identified with its *entropy*, which in turn is just the expected information (the sum of the information provided by each cell weighted by its probability)[25]:

(29) Entropy and Information: Let $Part(c) = \mathcal{Q} = \{q_1, \ldots, q_k\}$. Let $P(q_j)$ be the probability of q_j. Then:

a. Expected information: The entropy of $\mathcal{Q}$, $H(\mathcal{Q})$, is the expected information $H(\mathcal{Q}) = \sum_{j=1}^{k} P(q_j) inf(q_j)$[26].

[25] Shannon (1948) posited some basic axioms that any measure of relevant uncertainty should follow, and proved that (29) is the unique measure satisfying these axioms. Throughout this paper, we will assume that our logarithms are binary ($log_2 n$ is that number k such that $2^k = n$).

[26] To reduce clutter, we omit the multiplicative constant that is sometimes presented in the derivation of entropy.

b. Information: The information received from any particular cell q_j is $inf(q_j) = log_2(1/P(q_j))$.

(30) Examples:

a. Let $\mathcal{Q} = \{11, 10, 01, 00\}$, and suppose that the elements in $\mathcal{Q}$ have the same probability: $\forall q_j \in \mathcal{Q}: P(q_j) = 1/4$. Then for all $q_j \in \mathcal{Q}$, $inf(q_j) = log_2(4) = 2$, and $H(\mathcal{Q}) = 2$.

b. Let $\mathcal{Q} = \{11, 10, 01, 00\}$, and suppose that the elements in $\mathcal{Q}$ have the following probabilities: $P(11) = 1/8$, $P(10) = P(01) = 1/4$, $P(00) = 3/8$. Then $inf(11) = 0.53$, $inf(10) = inf(01) = 0.5$, $inf(00) = 0.375$, and thus $H(\mathcal{Q}) = 1.9$.

c. Let $\mathcal{Q} = \{111, 110, 101, 100, 011, 010, 001, 000\}$, and suppose that the elements in $\mathcal{Q}$ have the same probability: $\forall q_j \in \mathcal{Q}: P(q_j) = 1/8$. Then for all $q_j \in \mathcal{Q}$, $inf(q_j) = log_2(8) = 3$, and $H(\mathcal{Q}) = 3$.

The examples in (30) indicate some general properties that motivate the entropic measure of relevant uncertainty. When the elements of a partition $\mathcal{Q}$ have the same probability ($\forall q_i \in \mathcal{Q}: P(q_i) = 1/|\mathcal{Q}|$), the entropy is the log of the size of the set: $H(\mathcal{Q}) = log_2(|\mathcal{Q}|)$. This makes sense: each cell q_i provides information $log_2(1/P(q_i)) = log_2(1/(1/|\mathcal{Q}|)) = log_2(|\mathcal{Q}|)$, and since each cell is equally likely, $log_2(|\mathcal{Q}|)$ is the amount of information we expect to receive. Note also that the partition induced by considering whether k literals are true has entropy k when all cells are equally probable. Thus, when there are more literals, and hence more cells in the partition, there is more uncertainty. Finally, note that the entropy is reduced when probabilities are not equal (you are most uncertain when you have no bias among alternatives).

Assume now that the cost associated with relevant uncertainty in a context is identified with the information-theoretic entropy of the QUD in that context. Assume also (to keep calculations simple) that the cells in the partition have equal probability[27]. The logarithmic growth of entropy means that the corresponding cost functions are now more contained.

[27]This assumption might turn out to be problematic. It is conceivable that probabilities decrease with the number of true alternatives. For example, an *a priori* assumption that predicate extensions are as small as possible might provide a rationale for theories of "minimal worlds/models" theories of exhaustivity (e.g., van Benthem, 1989; van Rooij and Schulz, 2004; Spector, 2005, 2006, 2016; Schulz and van Rooij, 2006). Given the symmetry problem (von Fintel and Heim, 1999; see also Fox (2007), Katzir (2007)), the Maxim of Quantity cannot motivate the minimal worlds/models assumption. For example, suppose we learn from a speaker that $R(a)$ and we are in a context in which it is relevant whether b satisfies R. A speaker obeying the Maxim of Quantity could only be taken to be ignorant about whether $R(b)$. However, if $R(b)$ is *a priori* less likely than $\neg R(b)$, this might make it rational for the listener to conclude that $R(b)$ is false. More generally, it is plausible to assume that for an arbitrary predicate P and arbitrary individual c, $P(c)$ is less likely to be true than false. This assumption may relate to the "size principle" proposed in concept learning (e.g., Tenenbaum, 1999), and may also underlie our ability—granted by *exh*—to state only the positive instances of a predicate (these being the least likely, and hence worth the cost of expression). See also Bar-Hillel and Carnap (1952) on (a-)symmetries between a predicate and its negation. I hope to return to this set of ideas in future work.

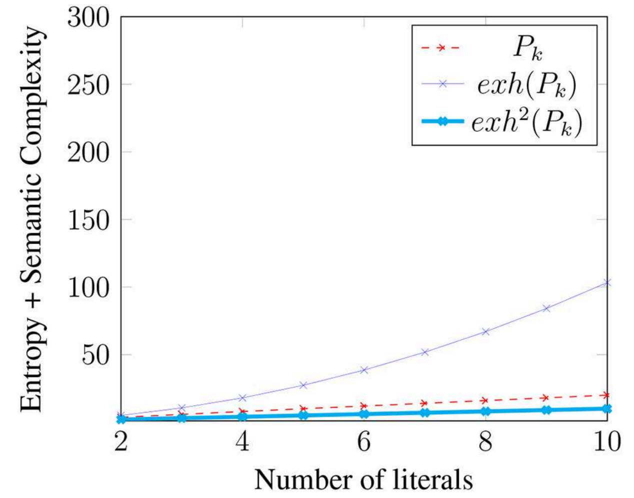


FIGURE 3 | Cost of update (final version).

(31) Costs of update (final version):

Formula	Complexity	Entropy
P_k	k	$log_2(2^k - 1)$
$exh(P_k)$	k^2	$log_2 k$
$exh^2(P_k)$	k	0

More to the point, we now predict the desired result that for all values of k, conjunctions are less costly than inclusive disjunctions which in turn are less costly than exclusive disjunctions (see **Figure 3**).

3.5. How Many Kinds of Cost?

With (31), we have completed our development of the cost function we stated in (19). We repeat the statement below in (32):

(32) Processing costs during disambiguation: Let S be a sentence uttered in context c. Suppose that grammar $\mathcal{G}$ assigns k analyses to S: $\mathcal{G}(S) = \{\lambda_1, \ldots, \lambda_k\}$, where each λ_i is a form-meaning pair $< f_i, m_i >$. Let $\mathcal{S}(m_i)$ be the semantic complexity of m_i, let c_i be the result of updating context c with m_i, $c + m_i$, and let $\mathcal{H}(c_i)$ be the entropy in context c_i. Then the *cost of selecting* $\lambda_i \in \mathcal{G}(S)$ in context c, $C(\lambda_i, c)$, is: $C(\lambda_i, c) = \mathcal{S}(m_i) + \mathcal{H}(c_i)$.

At first blush, the two kinds of cost seem different. Semantic complexity is a measure of compressibility: what is the smallest representation that can produce the desired meaning? Entropy is a measure of relevant uncertainty: how much information is needed to resolve our uncertainty? As it happens, entropy has a coding interpretation. Shannon (1948) noted that the entropy tells us the length of the representation (in bits) that would be needed to communicate outcomes in $\mathcal{Q}$[28]. Thus, both $\mathcal{S}$ and $\mathcal{H}$ give compression-based costs: semantic complexity tells us how much cost we have to pay for the current message, and entropy

[28]More generally, the *noiseless coding theorem* states that the minimal average code length for encoding outcomes in $\mathcal{Q}$ is very close to the entropy of $\mathcal{Q}$.

tells how much it would cost to get to a complete answer and hence how much cost we can expect to pay before our work is done[29].

We may also want a more general variant of (32) that allows for other kinds of costs to be incorporated, and for different ways of combining them. For example, it is natural to consider the possibility that the information of a given answer might itself have a cost, or that entropy reduction (the difference in entropy between the input and output contexts) is more central than the entropy in the output context alone. To allow for these and other possibilities in formulating theories of the cost function, a less committed variant would say that $C(\lambda_i, c)$ is a monotonically increasing function of $\mathcal{S}(m_i)$ and $\mathcal{H}(c_i)$.

4. CONCLUDING REMARKS

We have in (32) a function that assigns a cost to any given interpretation to an ambiguous sentence uttered in a context *c*. So far, I have said nothing about the disambiguation mechanism. I assume here that disambiguation decisions are made by finding optimal solutions to a coordination problem between speaker and hearer [see (6)]. In general, such decisions will involve assigning utilities to the space of output contexts, where coordination gets more utility than non-coordination and where the utilities might take the costs in (32) into account. There will also be a probability distribution over the space of output contexts (the probability that the speaker intends for each candidate to be the output context), and this will be partly determined by assumptions about the speaker's epistemic state. There will also be assumptions about what the QUD is, and these will determine (in conjunction with *exh* and the Maxim of Quantity) what the space of output contexts will be. In such a framework, the cost function puts a certain pressure to minimize costs (by the utility function), but the costs will be just one factor in the set of considerations that help a listener disambiguate. I should like to emphasize, however, that probabilities in this architecture only enter into disambiguation considerations, and hence the approach developed here is quite different than systems that allow probabilities to enter into the strengthening mechanism itself (e.g., Franke, 2011; Potts et al., 2015; Bergen et al., 2016). In the terminology of Fox and Katzir (2019), I assume that *exh* does not take a probability distribution as an argument, although the function that solves the decision problem in (6) does.

[29]Roni Katzir (p.c.) notes that the picture here is quite analogous to *Minimum Description Length* (MDL) approaches to learning (Rissanen, 1978). Such approaches compare competing hypotheses for a given set of data by minimizing the sum of (i) the cost to encode the hypothesis, and (ii) the cost to encode the data given the hypothesis. Semantic complexity straightforwardly relates to (i), but it is unclear (to me) how to relate entropy to (ii). For example, in MDL learning we compare hypotheses that can make sense of the data. In our disambiguation scenario, we have different form-meaning pairs that can be associated with the observable data (the sentence *S*), but what the entropy measure is concerned with is how these different analyses affect the context, and different analyses will (in general) lead to different contexts. I hope to return to the comparison with MDL in future work.

The cost function in (32) aims to make sense of CUPID, the puzzle of why and how exhaustification can be treated with uniformity in the competence system but with diversity in the performance system. I have argued that this can be made sense of by assuming that exhaustification itself is not the source of cost. Instead, I assume that costs are calculated by systems that ignore the computations internal to the language faculty. The cost calculation looks at the proposition denoted by each candidate analysis of the sentence, as well as the way this proposition would affect the information in the context, and assigns a cost to each using domain-general considerations. Like other models proposed from the early days of generative grammar (e.g., Miller and Chomsky, 1963) up to more modern treatments (e.g., Levy, 2013), my proposal here identifies a role for the complexity of the sentence itself as well as for information-theoretic reasoning about uncertainty resolution. However, the only aspect of the sentence that is relevant for our purposes is its meaning, with no regard for or access to its computational history.

The commitment to domain-general principles pursued here means that I have not considered language-dependent characterizations of scalar diversity in processing. For example, acquisition studies have argued that children differ from adults in one important way: they do not make lexical substitutions in generating *ALT* (e.g., Barner and Bachrach, 2010; Barner et al., 2011; Singh et al., 2016b; Tieu et al., 2017). One might pursue the idea that lexical substitutions, even when they emerge in the adult state, are the source of processing costs (see Chemla and Bott, 2014 and van Tiel and Schaeken, 2017 for steps in this direction). Note that free-choice inferences do not *require* lexical substitutions (the constituents are enough of a substitution source), and numerals cannot in general require lexical substitutions because the set of alternatives is infinite and hence must be generated by the successor function (see also section 1.4). This perspective would need to make sense of why lexical substitution does not seem to be hard with *only* (Marty and Chemla, 2013), and in any event working this all out raises non-trivial challenges that would take us too far afield to discuss here (Chemla and Singh, 2016). I hope to return to a fuller comparison in future work.

We have considered the idea that *exh* has several functions: it typically strengthens meanings, but it also often complicates meanings and gets us to better and better answers without having to verbalize them outright. Consider for example assertion of a disjunction $P_k = p_1 \vee p_2 \vee \ldots p_k$ in a world with no *exh* and in which the Maxim of Quantity governs communication. In such a world, you only eliminate one cell of the 2^k cells of the partition, and you thus generate lots of ignorance inferences (Fox, 2007). But suppose that the speaker in this world knows that exactly one of the p_i is true but doesn't know which. They would then have to produce a complex utterance to convey this thought: $(p_1 \wedge \neg p_2 \wedge \ldots \wedge \neg p_k) \vee (\neg p_1 \wedge p_2 \wedge \neg p_3 \wedge \ldots \neg p_k) \vee \ldots (\neg p_1 \wedge \ldots \wedge \neg p_{k-1} \wedge p_k)$. This is a k^2 mouthful. If a super-engineer were kind enough to give the speaker and hearer access to *exh*, they could communicate this complex piece of information by uttering P_k and hoping the listener would realize they should parse the sentence with *exh*. Presumably, the joint cost of *exh* and P_k, together with the risk of error (given the new ambiguity), is a better way to communicate a good and complex answer than having to utter it outright.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

My thinking on these topics has been shaped by discussions with many people. I would like in particular to thank Leon Bergen, Brian Buccola, Noam Chomsky, Luka Crnič, Kai von Fintel, Yossi Grodzinsky, Irene Heim, Philippe Schlenker, Jesse Snedeker, Benjamin Spector, Bob Stalnaker, Ida Toivonen, and especially Emmanuel Chemla, Roni Katzir, and Danny Fox. The paper benefitted from comments and questions received by two reviewers and the editors Sophie Repp and Katharina Spalek, as well as feedback received during presentations of some of this material at Carleton University, the Hebrew University of Jerusalem, MIT, and the University of Toronto. I would also like to acknowledge support from the Social Sciences and Humanities Research Council of Canada, Grant No. 435-2017-1256.

REFERENCES

van Benthem, J. (1989). "Semantic parallels in natural language and computation," in *Logic Colloquium '87*, eds H. -D. Ebbinghaus, J. Fernandez-Prida, M. Garrido, D. Lasscar, and M. Rodriguez Artalejo. (Amsterdam: Elsevier), 331–375.

van Benthem, J., Ebbinghaus, H.-D., Fernandez-Prida, J., Garrido, M., Lasscar, D., and Rodriguez Artalejo, M. (1986). *Essays in Logical Semantics*. Dordrecht: Reidel.

von Fintel, K. (2008). What is presupposition accommodation, again? *Philos. Perspect.* 22, 137–170. doi: 10.1111/j.1520-8583.2008.00144.x

von Fintel, K., and Heim, I. (1999). *Notes on Implicature*. Cambridge, MA: Lecture Notes, 24.954: Pragmatics in Linguistic Theory, MIT.

van Rooij, R. (2004). Utility, informativity, and protocols. *J. Philos. Logic* 33, 389–419. doi: 10.1023/B:LOGI.0000036830.62877.ee

van Rooij, R., and Schulz, K. (2004). Exhaustive interpretation of complex sentences. *J. Logic Lang. Inform.* 13, 491–519. doi: 10.1007/s10849-004-2118-6

van Tiel, B., and Schaeken, W. (2017). Processing conversational implicatures: alternatives and counterfactual reasoning. *Cognit. Sci.* 41, 1119–1154. doi: 10.1111/cogs.12362

van Tiel, B., van Miltenburg, E., Zevakhina, N., and Geurts, B. (2016). Scalar diversity. *J. Semant.* 33, 137–175.

Alonso-Ovalle, L. (2005). "Distributing the disjuncts over the modal space," in *Proceedings of NELS 35*, L. Bateman and C. Ussery (Amherst, MA: GLSA).

Bar-Hillel, Y., and Carnap, R. (1952). *An Outline of a Theory of Semantic Information*. Technical Report 247, Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, MA.

Bar-Lev, M., and Fox, D. (2017). "Universal free choice and innocent inclusion," in *Proceedings of SALT 27*.

Barner, D., and Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognit. Psychol.* 60, 40–62. doi: 10.1016/j.cogpsych.2009.06.002

Barner, D., Brooks, N., and Bale, A. (2011). Accessing the unsaid: the role of scalar alternatives in children's pragmatic inference. *Cognition* 118, 87–96. doi: 10.1016/j.cognition.2010.10.010

Beaver, D. (2001). *Presupposition and Assertion in Dynamic Semantics*. Stanford, CA: CSLI.

Bergen, L., Levy, R., and Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semant. Pragmat.* 9:20. doi: 10.3765/sp.9.20

Bott, L., Bailey, T. M., and Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *J. Mem. Lang.* 66, 123–142. doi: 10.1016/j.jml.2011.09.005

Bott, L., and Noveck, I. (2004). Some utterances are underinformative: the onset and time course of scalar inferences. *J. Mem. Lang.* 53, 437–457. doi: 10.1016/j.jml.2004.05.006

Bowler, M. (2014). "Conjunction and disjunction in a language without 'and'," in *Proceedings of SALT 24*, 137–155.

Breheny, R., Ferguson, H. J., and Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Lang. Cognit. Process.* 28, 443–467. doi: 10.1080/01690965.2011.649040

Breheny, R., Katsor, N., and Williams, J. (2006). Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100, 434–463. doi: 10.1016/j.cognition.2005.07.003

Buccola, B., Križ, M., and Chemla, E. (2018). *Conceptual Alternatives: Competition in Language and Beyond*. Paris: ENS.

Chemla, E. (2009a). *Similarity: Towards a Unified Account of Scalar Implicatures, Free Choice Permission, and Presupposition Projection*. Paris: ENS.

Chemla, E. (2009b). Universal implicatures and free choice effects: experimental data. *Semant. Pragmat.* 2, 1–33. doi: 10.3765/sp.2.2

Chemla, E., and Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: disjunctions and free choice. *Cognition* 130, 380–396. doi: 10.1016/j.cognition.2013.11.013

Chemla, E., Cummins, C., and Singh, R. (2016). Training and timing local scalar enrichments under global pragmatic pressures. *J. Semant.* 34, 107–126. doi: 10.1093/jos/ffw006

Chemla, E., and Singh, R. (2014a). Remarks on the experimental turn in the study of scalar implicature, Part I. *Lang. Linguist. Compass* 8, 373–386. doi: 10.1111/lnc3.12081

Chemla, E., and Singh, R. (2014b). Remarks on the experimental turn in the study of scalar implicature, Part II. *Lang. Linguist. Compass* 8, 387–399. doi: 10.1111/lnc3.12080

Chemla, E., and Singh, R. (2016). *Pruning, Shortcuts, and Sources of Complexity in the Computation of Exhaustivity*. Paris; Ottawa: ENS; Carleton University.

Chemla, E., and Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *J. Semant.* 28, 359–400. doi: 10.1093/jos/ffq023

Chierchia, G., Fox, D., and Spector, B. (2012). "Scalar implicature as a grammatical phenomenon," in *Handbook of Semantics*, Vol. 3, eds P. Portner, C. Maienborn, and K. Heusinger (New York, NY: Mouton de Gruyter), 2297–2331.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Crain, S., and Steedman, M. (1985). "On not being led up the garden path: the use of context by the psychological syntax processor," in *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, eds D. R. Dowty, L. Karttunen, and A. Zwicky (Cambridge: Cambridge University Press), 320–358.

Crnič, L., Chemla, E., and Fox, D. (2015). Scalar implicatures of embedded disjunctions. *Nat. Lang. Semant.* 23, 271–305. doi: 10.1007/s11050-015-9116-x

Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., and Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguist. Philos.* 21, 159–210. doi: 10.1023/A:1005330227480

Davidson, K. (2013). 'And' and 'or': general use coordination in ASL. *Semant. Pragmat.* 6, 1–44. doi: 10.3765/sp.6.4

De Neys, W., and Schaeken, W. (2007). When people are more logical under cognitive load: dual task impact on scalar implicature. *Exp. Psychol.* 54, 128–133. doi: 10.1027/1618-3169.54.2.128

Enguehard, E., and Chemla, E. (2019). *Connectedness as a Constraint on Exhaustification*. Paris: ENS.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature* 407, 630–633. doi: 10.1038/35036586

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A., Bever, T. G., and Garrett, M. F. (1974). *The Psychology of Language*. New York, NY: MgGraw-Hill.

Fox, D. (2007). "Free choice disjunction and the theory of scalar implicature," in *Presupposition and Implicature in Compositional Semantics*, eds U. Sauerland, and P. Stateva (New York, NY: Palgrave Macmillan), 71–120.

Fox, D. (2018). "Partition by exhaustification: comments on Dayal 1996," in *Proceedings of Sinn und Bedeutung*.

Fox, D., and Katzir, R. (2011). On the characterization of alternatives. *Nat. Lang. Semant.* 19, 87–107. doi: 10.1007/s11050-010-9065-3

Fox, D., and Katzir, R. (2019). *Modularity and Iterated Rationality Models of Scalar Implicatures*. Cambridge, MA; Tel Aviv: MIT; Tel Aviv University.

Fox, D., and Spector, B. (2018). Economy and embedded exhaustification. *Nat. Lang. Semant.* 26, 1–50. doi: 10.1007/s11050-017-9139-6 Accepted for publication in *Natural Language Semantics*.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semant. Pragmat.* 4, 1–82. doi: 10.3765/sp.4.1

Frazier, L. (1985). "Syntactic complexity," in *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, eds D. R. Dowty, L. Karttunen, and A. Zwicky (Cambridge: Cambridge University Press), 129–189.

Gajewski, J., and Sharvit, Y. (2012). In defense of the grammatical approach to scalar implicature. *Nat. Lang. Semant.* 20, 31–57. doi: 10.1007/s11050-011-9074-x

Grice, H. (1967). *Logic and Conversation*. Cambridge, MA: William James Lectures, Harvard University.

Groenendijk, J., and Stokhof, M. (1984). *Studies on the semantics of questions and the pragmatics of answers* (Doctoral dissertation), University of Amsterdam, Amsterdam, Netherlands.

Horn, L. R. (1972). *On the semantic properties of logical operators in English* (Doctoral dissertation), UCLA, Los Angeles, CA, United States.

Huang, Y. T., and Snedeker, J. (2009). Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. *Cognit. Pyschol.* 58, 376–415. doi: 10.1016/j.cogpsych.2008.09.001

Kamp, H. (1973). Free choice permission. *Arist. Soc. Proc.* 74, 57–74. doi: 10.1093/aristotelian/74.1.57

Katzir, R. (2007). Structurally defined alternatives. *Linguist. Philos.* 30, 669–690. doi: 10.1007/s10988-008-9029-y

Katzir, R. (2014). "On the role of markedness and contradiction in the use of alternatives," in *Semantics, Pragmatics, and the Case of Scalar Implicatures*, ed S. P. Reda (Basingstoke: Palgrave), 40–71.

Katzir, R., and Peled, N. (2018). *Representation and Learning of Quantificational Determiners*. Tel Aviv: Tel Aviv University.

Katzir, R., and Singh, R. (2013). Constraints on the lexicalization of logical operators. *Linguist. Philos.* 36, 1–29. doi: 10.1007/s10988-013-9130-8

Katzir, R., and Singh, R. (2015). "Economy of structure and information: oddness, questions, and answers," in *Proceedings of Sinn und Bedeutung 19*, eds E. Csipak, and H. Zeijlstra, 302–319.

Keenan, E. L., and Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguist. Philos.* 9, 253–326. doi: 10.1007/BF00630273

Kratzer, A., and Shimoyama, J. (2002). "Indeterminate pronouns: the view from Japanese," in *Proceedings of the 3rd Tokyo Conference on Psycholinguistics*, ed Y. Otsu, 1–25.

Levy, R. (2013). "Memory and surprisal in human sentence comprehension," in *Sentence Processing*, ed R. P. G. van Gompel (Hove: Psychology Press), 78–114.

Lewis, D. (1988). Relevant implication. *Theoria* 54, 161–174. doi: 10.1111/j.1755-2567.1988.tb00716.x

Magri, G. (2009). A theory of individual level predicates based on blind mandatory scalar implicatures. *Nat. Lang. Semant.* 17, 245–297. doi: 10.1007/s11050-009-9042-x

Magri, G. (2011). Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semant. Pragmat.* 4, 1–51. doi: 10.3765/sp.4.6

Marty, P., and Chemla, E. (2013). Scalar implicatures: working memory and a comparison with 'only'. *Front. Psychol.* 4:403. doi: 10.3389/fpsyg.2013.00403

Marty, P., Chemla, E., and Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua* 133, 152–163. doi: 10.1016/j.lingua.2013.03.006

Mascarenhas, S. (2014). *Formal semantics and the psychology of reasoning* (Doctoral dissertation), New York, NY: NYU.

McCluskey, E. J. (1956). Minimization of Boolean functions. *Bell Syst. Tech. J.* 35, 1417–1444. doi: 10.1002/j.1538-7305.1956.tb03835.x

Meyer, M.-C. (2015). Generalized free choice and missing alternatives. *J. Semant.* 33, 703–754. doi: 10.1093/jos/ffv010

Meyer, M.-C., and Sauerland, U. (2009). A pragmatic constraint on ambiguity detection. *Nat. Lang. Linguist. Theory* 27, 139–150. doi: 10.1007/s11049-008-9060-2

Miller, G., and Chomsky, N. (1963). "Finitary models of language users," in *Handbook of Mathematical Psychology*, Vol. 2, eds R. Luce, R. Bush, and E. Galanter (New York, NY: Wiley), 419–491.

Noveck, I. (2001). When children are more logical than adults: Investigations of scalar implicature. *Cognition* 78, 165–188. doi: 10.1016/S0010-0277(00)00114-1

Piantadosi, S. T., Tenenbaum, J. B., and Goodman, N. D. (2016). The logical primitives of thought: empirical foundations for compositional cognitive models. *Psychol. Rev.* 123, 392–424. doi: 10.1037/a0039980

Podlesny, O. (2015). *Investigating disjunction in American Sign Language: the importance of nonmanual signals and the influence of English* (Master's thesis), Carleton University, Institute of Cognitive Science, Ottawa, ON, Canada.

Potts, C., Lassiter, D., Levy, R., and Frank, M. C. (2015). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *J. Semant.* 33, 755–802. doi: 10.1093/jos/ffv012

Quine, W. (1952). The problem of simplifying truth functions. *Am. Math. Monthly* 59, 521–531. doi: 10.1080/00029890.1952.11988183

Quine, W. (1955). A way to simplify truth functions. *Am. Math. Monthly* 62, 627–631. doi: 10.1080/00029890.1955.11988710

Rissanen, J. (1978). Modeling data by shortest description. *Automatica* 14, 465–471. doi: 10.1016/0005-1098(78)90005-5

Savage, J. E. (1976). *The Complexity of Computing*. New York, NY: Wiley.

Schulz, K., and van Rooij, R. (2006). Pragmatic meaning and non-monotonic reasoning: the case of exhaustive interpretation. *Linguist. Philos.* 29, 205–250. doi: 10.1007/s10988-005-3760-4

Shannon, C. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Singh, R. (2008a). *Modularity and locality in interpretation* (Doctoral dissertation), Cambridge, MA: MIT.

Singh, R. (2008b). On the interpretation of disjunction: asymmetric, incremental, and eager for inconsistency. *Linguist. Philos.* 31, 245–260. doi: 10.1007/s10988-008-9038-x

Singh, R., Fedorenko, E., Mahowald, K., and Gibson, E. (2016a). Accommodating presuppositions is inappropriate in implausible contexts. *Cognit. Sci.* 40, 607–634. doi: 10.1111/cogs.12260

Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., and Fox, D. (2016b). Children interpret disjunction as conjunction: consequences for theories of implicature and child development. *Nat. Lang. Semant.* 24, 305–352. doi: 10.1007/s11050-016-9126-3

Sipser, M. (1997). *Introduction to the Theory of Computation*. Boston, MA: PWS Publishing Company.

Spector, B. (2005). "Scalar implicatures: exhaustivity and Gricean reasoning," in *Questions in Dynamic Semantics*, eds M. Aloni, A. Butler, and P. Dekker (Amsterdam: Elsevier).

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques* (Doctoral dissertation), Paris: Université Paris 7.

Spector, B. (2013). Bare numerals and scalar implicatures. *Lang. Linguist. Compass* 7, 273–294. doi: 10.1111/lnc3.12018

Spector, B. (2016). Comparing exhaustivity operators. *Semant. Pragmat.* 9, 1–33. doi: 10.3765/sp.9.11

Stalnaker, R. (1978). "Assertion," in *Pragmatics*, ed P. Cole (New York, NY: Academic Press), 315–332.

Szymanik, J., and Thorne, C. (2017). Exporing the relation between semantic complexity and quantifier distribution in large corpora. *Lang. Sci.* 60, 80–93. doi: 10.1016/j.langsci.2017.01.006

Tenenbaum, J. B. (1999). *A Bayesian framework for concept learning* (PhD thesis), MIT, Cambridge, MA, United States.

Tieu, L., Romoli, J., Zhou, P., and Crain, S. (2016). Children's knowledge of free choice inferences and scalar implicatures. *J. Semant.* 33, 269–298. doi: 10.1093/jos/ffv001

Tieu, L., Yatsuhiro, K., Cremers, A., Romoli, J., Sauerland, U., and Chemla, E. (2017). On the role of alternatives in the acquisition of simple and complex disjunctions in French and Japanese. *J. Semant.* 34, 127–152. doi: 10.1093/jos/ffw010

Trakhtenbrot, B. A. (1984). A survey of Russian approaches to *Perebor* (brute-force search) algorithms. *Ann. Hist. Comput.* 6, 384–400. doi: 10.1109/MAHC.1984.10036

Williamson, T. (1996). Knowing and asserting. *Philos. Rev.* 105, 489–523. doi: 10.2307/2998423

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 06 November 2019
doi: 10.3389/fpsyg.2019.02329

Children's Acquisition of Homogeneity in Plural Definite Descriptions

Lyn Tieu[1]*, *Manuel Križ*[2] *and Emmanuel Chemla*[3]

[1] Office of the Pro Vice-Chancellor (Research and Innovation), School of Education, MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, NSW, Australia, [2] Department of Linguistics, University of Vienna, Vienna, Austria, [3] Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, ENS, PSL University, EHESS, CNRS, Paris, France

Plural definite descriptions give rise to *homogeneity* effects: the positive *The trucks are blue* and the negative *The trucks aren't blue* are both neither true nor false when some of the trucks are blue and some are not, that is, when the group of trucks is not *homogeneous* with respect to the property of being blue (Löbner, 1987, 2000; Schwarzschild, 1994; Križ, 2015b). The only existing acquisition studies related to the phenomenon have examined children's comprehension only of the affirmative versions of such sentences, and moreover have yielded conflicting data; while one study reports that preschoolers interpret definite plurals *maximally* (Munn et al., 2006, see also Royle et al., 2018), two other studies report that preschoolers allow *non-maximal* interpretations of definite plurals where adults do not (Karmiloff-Smith, 1979; Caponigro et al., 2012). Moreover, there is no agreed upon developmental trajectory to adult homogeneity. In this paper, we turn to acquisition data to investigate the predictions of a recent analysis of homogeneity that treats homogeneous meanings as the result of a scalar implicature (Magri, 2014). We conducted two experiments targeting 4- and 5-year-old French-speaking children's interpretations of plural definite descriptions in positive and negative sentences, and tested the same children on standard cases of scalar implicature. The experiments revealed three distinct subgroups of children: those who interpreted the plural definite descriptions existentially and failed to compute implicatures; those who both accessed homogeneous interpretations and computed implicatures; and finally, a smaller subgroup of children who appeared to access homogeneous interpretations without computing implicatures. We discuss the implications of our findings, which appear to speak against the implicature theory as the adult-like means of generating homogeneous meanings.

Keywords: homogeneity, language acquisition, alternatives, scalar implicature, definite descriptions, quantification, plurals, maximality

OPEN ACCESS

Edited by:
Katharina Spalek,
Humboldt University of Berlin, Germany

Reviewed by:
Sebastian Löbner,
Heinrich Heine University of Düsseldorf, Germany
Jill Gibson De Villiers,
Smith College, United States

Correspondence:
Lyn Tieu
lyn.tieu@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *02 April 2019*
Accepted: *30 September 2019*
Published: *06 November 2019*

Citation:
Tieu L, Križ M and Chemla E (2019) Children's Acquisition of Homogeneity in Plural Definite Descriptions. Front. Psychol. 10:2329. doi: 10.3389/fpsyg.2019.02329

1. INTRODUCTION

Plural definite descriptions give rise to *homogeneity* effects (see among others, Fodor, 1970; Schwarzschild, 1994; Löbner, 2000; Breheny, 2005; Gajewski, 2005; Büring and Križ, 2013; Spector, 2013; Magri, 2014; Križ, 2015a). The positive (1) is true in a situation where all of the trucks are blue, but its negation (2) is only true in a situation where none of them are. There is a gap, however,

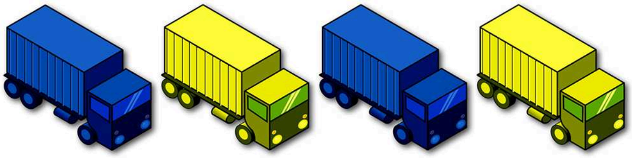


FIGURE 1 | Image corresponding to a GAP context. The first and third trucks are blue, while the second and fourth trucks are yellow.

in between these two possibilities; in a situation where some but not all of the trucks are blue (**Figure 1**), neither the positive sentence nor its negation are true. In this particular GAP context, the group of trucks is not *homogeneous* with respect to the property of being blue[1].

(1) The trucks are blue.

(2) The trucks are not blue.

Now compare (1) and (2) to the universally quantified (3) and (4). At first glance, the positive (1) might appear to be interpreted roughly equivalently to the universally quantified (3). Yet this apparent equivalency between *the*-NP and *all*-NP disappears under negation: in contrast to (2), the negative (4) is true in the scenario depicted in **Figure 1**.

(3) All of the trucks are blue.

(4) Not all of the trucks are blue.

The sentences with universal descriptions have complementary negations: the set of situations in which the positive sentence is true is the complement of the set of situations in which its negation is true, with no gap between them.

One way of conceptualizing the state of affairs for the definite descriptions is to say that in a GAP scenario, both the positive and negative sentences are neither true nor false; rather they correspond to a third truth value, or to none at all. Some experimental evidence for this can be found in a study by Križ and Chemla (2015), who presented adults with such sentences as descriptions of situations that violated homogeneity. They reported that adults often assessed such descriptions as neither completely true nor completely false. In contrast, sentences containing universal descriptions like (3) did not display this gap, and were simply judged as completely false in the same non-homogeneous scenarios.

In the present paper, we investigate the acquisition of such *truth value gaps*. Building on Križ and Chemla (2015), we will take the pattern they observed in adults as the empirical hallmark of homogeneity: their adult participants assessed positive definite descriptions and their negations as non-true in GAP contexts[2]. Now, if young children do not initially display this hallmark of homogeneity, one might expect them instead to assign complementary truth conditions to the positive and negative counterparts. In particular, one might expect children to liken plural definite descriptions (5) to existential quantifiers (6) or to universal quantifiers (7).

(5) a. The trucks are blue.
b. The trucks aren't blue.

(6) a. There are some blue trucks.
b. There aren't any blue trucks.

(7) a. Every truck is blue.
b. Not every truck is blue.

A child who is presented with (5) in a GAP context like **Figure 1**, then, might be expected to respond in one of three ways, depending on the interpretation assigned to the plural definite. First, if the child is adult-like, she can be expected to treat the positive and negative descriptions uniformly, likely rejecting both as descriptions of GAP contexts. This possibility corresponds to the HOMOGENEOUS pattern depicted in **Figure 2**. Second, the child could interpret the definite existentially, i.e., in parallel with (6), prompting her to accept the positive sentence but not the negative sentence as a good description of **Figure 1**. This corresponds to the EXISTENTIAL pattern depicted in **Figure 2**. Third, the child could interpret the definite universally, in line with (7), prompting her to accept the negative but not the positive description, as in the UNIVERSAL pattern in **Figure 2**[3].

To our knowledge, there are only three existing studies that have specifically investigated children's comprehension of plural definite descriptions, examining in particular whether children assign *maximal* interpretations to plural definite descriptions. Karmiloff-Smith (1979) and Caponigro et al. (2012) report corroborating findings that children allow *non-maximal* interpretations of positive plural definite descriptions where adults would not. Such children would be expected to accept a sentence like (1) as a description of a context like **Figure 1**. While these previous experiments leave open the status of negative

[1]Note that homogeneity effects appear not only with definite plurals, but also when a predicate is ascribed to a single object that has parts to which the predicate is applicable: (ia) is true if the entire truck is blue, and (ib) is true if no appreciable part of the truck is blue.

(i) a. The truck is blue.
b. The truck isn't blue.

In this paper, however, we restrict our attention to the case of definite plurals.

[2]Here and elsewhere in the paper, we will sometimes sloppily use the term *definite description* to refer to a sentence containing a definite description in subject position.

[3]This pattern seems particularly plausible when one considers the kinds of contexts in which children might hear the plural definite being uttered. If indeed the positive (1) is only felicitous and true in contexts where all of the trucks are blue, the child should only ever hear such descriptions in scenarios that satisfy homogeneity. This could lead the child to form the generalization that plural definites have a universal meaning. Some empirical evidence for this state of affairs in the caregiver input may be found in corpus data reported by Caponigro et al. (2012). These authors examined child-directed speech in the CHILDES database (MacWhinney, 2000), and reported that all 6404 instances in their sample of plural definite descriptions headed by the determiner *the* referred to a maximal element. Note that evidence about negative plural definites may be more intricate, for reasons having to do with scope; if a child only ever hears the negative (2) in homogeneous contexts, the child might be led to liken the plural definite to a universal that must take wide scope with respect to negation.

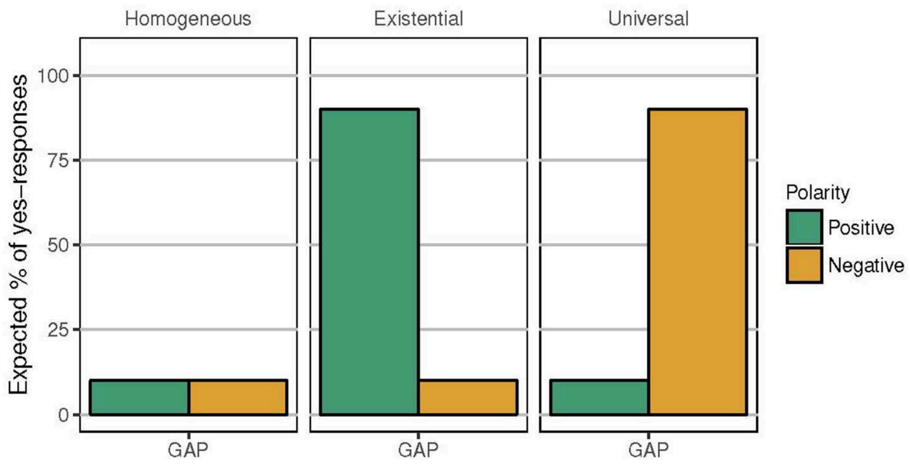


FIGURE 2 | Expected response patterns for positive and negative definite descriptions such as *The trucks are blue* and *The trucks aren't blue* in GAP contexts such as **Figure 1**, according to the interpretation of the plural definite description.

descriptions like (2), a problem that we will address shortly, they nevertheless provide a preliminary suggestion that children may not be sensitive to the truth value gap described above.

Karmiloff-Smith (1979) and Caponigro et al. (2012) provide different characterizations of their child participants' failure to enforce maximality in their interpretations of plural definite descriptions, although neither characterization provides an explanation of how children acquire maximality, nor of why it emerges relatively late (reportedly after 6 years of age). To date, there exists no unified explanation for these previous findings. Since these two studies were conducted, however, a recent semantic analysis of homogeneity has emerged which invokes a connection between the homogeneity that is triggered by plural definite descriptions and the enrichment mechanism that underlies the derivation of scalar implicatures (Magri, 2014). In what follows, we will investigate the precise predictions that such an analysis makes for children's development both of homogeneity and scalar implicatures, through two novel experiments.

The remainder of this paper is organized as follows. We will begin by briefly outlining the existing analyses of homogeneity in the semantics literature. We will then review the existing acquisition studies of plural definite descriptions, which raise as-of-yet unanswered questions about children's early interpretations of plural definite descriptions and about the learnability of homogeneity more generally. Since the scalar implicature account makes concrete predictions that one can test, we proceed to present two experiments where we did just that. We then discuss the implications of our findings for a theory of homogeneity and for the developmental trajectory toward adult homogeneity[4].

[4]A reviewer points out that the current study might also be seen as a contribution to the broader investigation of the acquisition of predication, truth-value gaps, and negation more generally, that is, above and beyond definite plurals. In what follows, however, we will focus our attention specifically on how the child data can be brought to bear on the implicature approach to the phenomenon.

2. THEORIES OF HOMOGENEITY

A few accounts of homogeneity have been proposed in the formal semantics literature. The earliest proposals treat homogeneity as a presupposition (Schwarzschild, 1994; Löbner, 2000; Gajewski, 2005). The general idea is that sentences like (1) and (2) carry a presupposition that either all of the trucks are blue or none of the trucks are blue. Since this presupposition is not satisfied in GAP contexts like **Figure 1**, such descriptions give rise to a truth value gap.

A second approach is to say that there is some sort of indeterminacy or vagueness about the interpretation of the definite description, which itself might be either existential or universal. A sentence is then perceived as having a definite truth value if it has the same truth value no matter how this indeterminacy is resolved (Spector, 2013; Križ and Spector, 2017). For example, if *the trucks* in (1) can be interpreted either existentially or universally, we have two possible interpretations for the sentence:

(8) a. Some of the trucks are blue.
b. All of the trucks are blue.

The sentence in (1) is then true if both (8a) and (8b) are true, i.e., if all of the trucks are blue, and false if both (8a) and (8b) are false, i.e., if none of the trucks are blue. In **Figure 1**, neither condition is satisfied, and so (1) can be neither true nor false. The same reasoning applies to the negative sentence (2), since the negations of (8a) and (8b) are neither both true nor both false.

A third approach derives homogeneity as a scalar implicature. Magri (2014) proposes that plural definites have a literal existential meaning that is strengthened to the universal meaning through an implicature[5]. Take the example of the scalar implicature in (9).

[5]We do not discuss here the very recent work by Bar-Lev (2018), which post-dates the writing of this paper. Bar-Lev proposes to view homogeneity as an implicature(-like) phenomenon in a way altogether different from Magri; the experimental data we present will not bear directly on Bar-Lev's approach.

(9) a. Some of the trucks are blue.
b. ⇝ *Not all of the trucks are blue*

The implicature in (9b) arises as the consequence of comparing the assertion in (9a) with alternatives that could have been uttered but were not. Assuming speakers are as informative as they can be (Grice, 1975), the speaker's choice to utter (9a), as opposed to the stronger alternative *All of the trucks are blue*, can lead us to conclude that this stronger alternative is false, generating the scalar implicature in (9b). This process by which the scalar implicature is derived can be analyzed as involving a covert exhaustification operator EXH, roughly equivalent to a silent "only" (Fox, 2007; Chierchia et al., 2011):

(10) EXH(Some of the trucks are blue)
= Some of the trucks are blue and not all of the trucks are blue

According to Magri (2014), homogeneity can be derived by recursively applying this exhaustification procedure. Assume first that the definite plural *the trucks* has a plain existential meaning, much like *some trucks* in (9a). Assume further that the lexical alternatives for the definite include "some" (though crucially not "all"). Now if we apply the same exhaustification procedure as in (10), but do so recursively, we effectively arrive at a universal meaning for (1), as in (11).

(11) EXH(EXH(The trucks are blue))
= EXH(The trucks are blue) and NOT(EXH(some of the trucks are blue))
= Some of the trucks are blue and NOT(some but not all of the trucks are blue)
= All of the trucks are blue

Of the three accounts outlined above, the scalar implicature account of homogeneity is of particular interest from a developmental perspective, in part because there exists a considerable amount of previous literature on the acquisition of scalar implicatures. This previous work will afford us a convenient means to empirically compare the two phenomena in development, and in doing so, to test the predictions of the theory[6].

2.1. Testing the Predictions of the Implicature Account

An implicature account of homogeneity *prima facie* predicts that children should perform on homogeneity the way that they perform on implicatures. After all, the same mechanism would underlie the strengthened meaning of a scalar term like "some" and the strengthened homogeneous meaning of a plural definite description.

There have been a number of developmental studies focusing on implicatures. Many of the existing studies have reported that children typically compute fewer scalar implicatures than adults (see among many others, Braine and Rumain, 1981; Chierchia et al., 2001; Gualmini et al., 2001; Noveck, 2001; Papafragou and Musolino, 2003; Barner et al., 2011). More recent developmental work on implicatures has shown that children's success on implicatures can in fact vary considerably, depending on factors such as the methodology being used to test the child's knowledge of implicature, the particular scale being tested, and the kinds of experimental contexts in which the scalar items are presented. For example, Katsos and Bishop (2011) have shown that providing 5-year-old children with three graded response options vastly improves the children's performance on implicatures, compared to when they are presented with the more traditional binary yes/no response options. That is, when children are given the option to reward a puppet with a minimal, intermediate, or maximal reward, they tend to perform in more of an adult-like manner, offering the intermediate reward for literally true but underinformative statements. Katsos and Bishop (2011) propose that children are simply more *pragmatically tolerant* than adults are when forced to decide whether or not to accept an underinformative statement.

Another proposal that has gained traction in the developmental literature is the idea that children's performance on implicatures is somehow linked to the nature of the *alternatives* that are involved in computing the implicature, with potential difficulties arising from accessing lexical alternatives or understanding their relevance in a given context (Barner et al., 2011; Singh et al., 2016; Skordos and Papafragou, 2016; Tieu et al., 2016). In particular, children appear to exhibit greater difficulties with implicatures that involve *lexical replacement* of alternative scalar terms, e.g., *some*/*all*, *or*/*and*, and *might*/*must*. By contrast, children have been reported to successfully compute *ad hoc* implicatures ("My friend has glasses" ⇝ *My friend doesn't have both glasses and a hat*) (Stiller et al., 2015) and free choice inferences ("Kungfu Panda may push the green car or the red car" ⇝ *Kungfu Panda may push the green car and Kungfu Panda may push the red car*) (Tieu et al., 2016), as well as conjunctive inferences from disjunction ("The chicken pushed a bus or an airplane" ⇝ *The chicken pushed a bus and an airplane*) (Singh et al., 2016; Tieu et al., 2017). These inferences share a common property: they do not involve lexical replacement; rather, children can retrieve the required alternatives directly from the test sentences or from the experimental context.

Given the insights of these recent studies on implicatures, we will set out to test the implicature account of homogeneity in a carefully controlled, systematic way, keeping in mind the role that alternatives, methodology, and context can play. We will systematically compare homogeneity to an implicature that, on the implicature theory, actually corresponds to a *sub-computation* of the homogeneity implicature [recall that (10) is a sub-computation of (11)]. Importantly, we will also use exactly the same tasks and contexts to test the two phenomena. This means that whatever effect the context may have on the one, it should also have the same effect on the other. Moreover, because the lexical alternatives involved in generating the *not-all* implicature and the homogeneity implicature are the very same, i.e., "some" and "all", we do not have to worry that children

[6]We have chosen to focus on the scalar implicature account primarily for practical reasons, as there is more existing work on the acquisition of implicatures than on the acquisition of vagueness or presupposition. We leave for future research an investigation of the predictions that alternative accounts of homogeneity may make for child language.

may acquire the alternatives for one inference earlier than for the other. In fact, the implicature theory in this case makes very straightforward, testable predictions.

If homogeneity is derived using the same mechanism as classical scalar implicatures, one should expect children to display sensitivity to homogeneity only once they are able to compute scalar implicatures, and more specifically only once they are able to compute the *not-all* implicature, since this corresponds to a sub-computation of the implicature of homogeneity. Previous studies have shown that without special training or facilitation, preschoolers typically respond to "some" statements in a manner consistent with the literal existential interpretation of the quantifier (e.g., Papafragou and Musolino, 2003). The implicature theory therefore predicts a similar pattern for homogeneity for such children, namely literal, existential interpretations of plural definite descriptions. Only once the children are capable of computing the *not-all* implicature will they display homogeneity effects. Furthermore, the implicature that gives rise to homogeneity effects involves *recursive* application of the exhaustification operator, so one might expect to see homogeneity surface even later in development than the regular first-order *not-all* implicature. While there is independent evidence that children are capable of recursive exhaustification (Zhou et al., 2013; Tieu et al., 2016), crucially, the timing prediction remains the same: we should not observe homogeneity surfacing *before* the scalar implicature.

3. ACQUISITION OF HOMOGENEITY

While there are existing studies of the acquisition of definite noun phrases on the one hand (see among others, Maratsos, 1974, 1976; Schafer and de Villiers, 2000; Matthewson et al., 2001; Pérez-Leroux et al., 2004; Schaeffer and Matthewson, 2005; Schmerse et al., 2014), and of plurality on the other hand (e.g., Berko, 1958; Winitz et al., 1981; Mervis and Johnson, 1991; Marcus et al., 1992; Fenson et al., 1994; Marchman et al., 1997; Sauerland et al., 2005; Barner et al., 2007; Zapf and Smith, 2008; Wood et al., 2009; Tieu et al., 2014; Davies et al., 2016), few studies have examined the two phenomena in conjunction. To our knowledge, there are only three existing studies that have specifically investigated children's comprehension of plural definite descriptions.

3.1. Karmiloff-Smith (1979)

In the earliest of these studies, Karmiloff-Smith (1979) reports a series of experiments investigating French-speaking children's production and comprehension of different kinds of noun phrases, including definite plural noun phrases. Two of Karmiloff-Smith's studies are relevant for our purposes here. First, she conducted a production study in which a child was prompted to produce directives, such as *Il faut mettre les camions dans le garage* "One must put the trucks in the garage." The experimental set-up involved two experimenters. One experimenter (E2) would turn his back and close his eyes. The other experimenter (E1) would manipulate a series of objects, for example, moving a set of toy trucks into the garage. The child would then have to tell E2 what he would have to do to replicate that action. E1 would return the objects to their original locations, and E2 would then turn around and open his eyes, and carry out the action based on the child's directive. By manipulating what sets of objects were moved into the garage, the experimenters aimed to elicit different kinds of noun phrases from the child, e.g., *les camions* "the trucks," *mes camions* "my trucks," *les camions bleus* "the blue trucks," etc. The experimenters tested children between the ages of 4;07 and 11;05. Karmiloff-Smith reports that for 4- and 5-year-olds, the definite article *les* was used to mark pluralization but not "totalization"; that is, *les X* was taken to signify any plural amount of X's, though not necessarily all the X's.

In a comprehension experiment modeled similarly to the production experiment, children were on the receiving end of the directives, and had to manipulate toy objects in response to these directives. For instance, they would hear sentences such as *Mets les voitures au garage* "Put the cars in the garage." Karmiloff-Smith reports low percentages of correct responses from 4- and 5-year-olds, suggesting again that the definite *les X* for these children signified any plurality of X's, though not necessarily the full set of X's. More generally, Karmiloff-Smith proposes that children initially mark newly acquired functions, such as pluralization, or totalization, through separate morphemes. In the earliest stage, between 3 and 5.5 years of age, children associate the plural definite *les* only with pluralization. In a second stage, between 5 and 8 years of age, children add the universal marker *tous* 'all' to convey totalization. Finally, after the age of 8 years, the definite plural *les* comes to simultaneously convey pluralization and totalization.

While Karmiloff-Smith's (1979) data are suggestive of what we have referred to in section 1 as the EXISTENTIAL pattern of interpretation, notice that the experiments she reports did not include plural definite descriptions under negation. The study therefore leaves open the status of children's interpretation of the negations of such plural definite descriptions, and does not allow us to fully determine which of the three scenarios in **Figure 2** the child's initial state corresponds to.

3.2. Munn et al. (2006)

The second study that has examined children's understanding of plural definite descriptions is reported in Munn et al. (2006). These authors compared children's understanding of singular and plural definite descriptions and indefinite nominals in English and Spanish. Like Karmiloff-Smith's comprehension study, Munn et al.'s study employed an act-out task. Preschoolers (mean age 4;01) were issued requests, such as "give me *the frogs next to the barn*", where there was a set of toy frogs beside a toy barn. The authors report that almost all children gave the maximal element of the relevant set of frogs 95% of the time[7]. In contrast to the conclusions reached in Karmiloff-Smith (1979),

[7]Royle et al. (2018) report a replication of Munn et al.'s study in French, using the same act-out task to test (Canadian) French-speaking children's comprehension of requests such as: *Donne-moi les vaches qui sont à côté de la ferme* "Give me the cows that are beside the farm." Children and adults generally gave maximal responses to definite plurals, with a significant effect of age observed for the children: the youngest children provided ∼37% non-maximal responses while older children gave ∼10% non-maximal responses (Royle et al., 2018:7).

the authors conclude that children correctly interpret plural definite descriptions maximally by the age of 3 years.

As pointed out in later work by Caponigro et al. (2010, 2012), however, there are some issues with this study. First, there were no control trials involving descriptions such as *some of the frogs*, so it is not clear whether children would also select the maximal set for such requests. Second, Caponigro et al. (2010) point out that Munn et al.'s reported percentage of maximal responses was calculated after excluding children who gave only one item in response to the plural definite description request; these children were clearly not assigning a maximal interpretation to the plural definite description. Third, Caponigro et al. (2012) point out that since Munn et al. (2006) did not provide a breakdown of the data by age, it is difficult to draw conclusions about when maximality in plural definite descriptions is acquired. Finally, like Karmiloff-Smith's (1979) study, this study, too, leaves open the status of children's interpretation of the negative definite description counterparts, without which we cannot tell whether the reported "maximal" behavior is due to an adult (homogeneous) interpretation of the plural definite description, or merely to a universal interpretation of the definite description.

3.3. Caponigro et al. (2012)

Caponigro et al. (2010, 2012) set out to investigate the possible developmental connection between plural definite descriptions like *the things on the plate* and free relative clauses like *what is on the plate*. The authors first conducted a Truth Value Judgment Task (TVJT) (Crain and Thornton, 1998) with 4-, 5-, 6-, and 7-year-old children, and a group of adult controls. In this task, participants were introduced to a character (Cookie Monster) who loves cookies but strongly dislikes onions. On critical target trials, children were presented with a picture of a plate containing three cookies and three onions, and were asked questions such as "Does Cookie Monster like *the things on the plate*?" or "Does Cookie Monster like *what's on the plate*?" The authors report that overall, free relatives and plural definite descriptions were interpreted maximally more frequently than existential nominals containing *one* and *some*, but less frequently than those containing the universal *all*[8]. As the authors point out, there are a couple of reasons to pursue the investigation further. First, even the adult controls that they tested did not always access maximal readings for the plural definite descriptions and free relatives, making it difficult to assess children's performance on the task. The authors suggest the problem may lie in the nature of the TVJT; they reason that if the plural definite descriptions introduced a presupposition of homogeneity, this presupposition was necessarily violated on the critical "mixed plate" trials, and so there could be no true or false answer given to the critical test questions. A second issue that the authors point out is that up until age 7, participants' responses to the critical trials were not different from chance; it is therefore unclear whether participants were simply guessing at random. Finally, as we pointed out previously for Munn et al.'s study, a maximal answer to positive sentences could be obtained either through homogeneity, or through a mere universal (non-adult-like) interpretation of the plural definite description.

To address the potential felicity issue with the use of the TVJT, the authors next conducted an act-out task, again with 4-, 5-, 6-, and 7-year-olds, and adult controls. In this task, participants were issued requests such as "Can you give me *the things on the plate*?" and "Can you give me *what's on the plate*?" The authors also compared the target conditions with ones in which the request contained *some*, *all*, and the nonsense determiner *blick*. Two of their main findings are relevant for us here. First, the authors reported a significant main effect of Question Type, with plural definite description responses differing from those in the *some* and *all* conditions. Second, further analysis revealed that the responses of the 4- and 5-year-olds, but not those of the 6- and 7-year-olds, were significantly different from those of adults; crucially, 4- and 5-year-olds assigned fewer maximal interpretations to the plural definite descriptions than the older children or the adults.

Caponigro et al. propose that although young children are capable of representing plural individuals, they struggle to map the conceptual/semantic representations of plural individuals to the relevant linguistic structure. These authors assume that the definite determiner denotes a function that applies to a set of individuals and returns the maximal element of that set (Link, 1983). They propose that young children associate the plural noun phrase with a set containing a plurality of atomic individuals, but one that contains no plural individuals or maximal individual. The meaning of *the* cannot apply to a set lacking a maximal individual, and so the semantic derivation fails, leading to the absence of maximal interpretations. The authors suggest that 4-year-olds must adopt other (possibly non-grammatical) strategies to deal with this failure, and point to the fact that their 4-year-olds treated the plural definite descriptions the same as they did the nonsense determiner *blick*.

3.4. Taking Stock

The previous acquisition studies described above tackled the question of whether young children enforce maximal interpretations on plural definite descriptions. The findings of Karmiloff-Smith (1979) and Caponigro et al. (2012) align, revealing that both French- and English-speaking children fail to interpret plural definite descriptions maximally until at least 6 years of age.

The findings of both of these studies raise three important questions. First, what underlies young children's non-maximal interpretations of plural definite descriptions? Second, what is the developmental trajectory that children take toward maximal interpretations? Finally, what triggers maximal interpretations, and so late in development? The two studies that report non-maximal behavior do not readily provide an answer to the third question, nor do they agree on the answers to the first two questions. On Karmiloff-Smith's proposal, children in the earliest stages associate the plural definite description with plurality, and not maximality. Children subsequently develop knowledge of the totalization function, and only later allow the plurifunctional/simultaneous marking of pluralization and totalization through the same morpheme. On Caponigro

[8]See Modyanova and Wexler (2008) for further evidence of non-maximal interpretations of free relatives.

et al.'s proposal, children initially fail to access maximal interpretations because they associate the plural noun phrase with a set of plural atomic individuals that lacks a maximal individual. What is missing, the authors speculate, is an adult-like mapping between the target linguistic structure and the relevant conceptual representation.

The finding that young children as a group do not interpret plural definite descriptions as maximally as adults do, does not rule out the possibility that they nevertheless interpret these expressions in systematic ways, and in particular, in a manner consistent with one of the possibilities presented in **Figure 2**. Unfortunately, none of the previous studies allow us to determine which scenario in **Figure 2** young children fall into, since these studies did not examine plural definite descriptions under negation[9]. Moreover, the data from these previous studies hint at more than one possibility. Specifically, Karmiloff-Smith's and Caponigro et al.'s participants who gave non-maximal responses could conceivably have assigned an existential interpretation to the plural definite description; Munn et al.'s participants, who gave maximal responses, could have interpreted the definite plural either universally or homogeneously. The first goal of our study, then, is to resolve this uncertainty surrounding the interpretations children assign to plural definite descriptions. In order to do so, we will examine children's interpretation of plural definite descriptions in both positive (upward-entailing) and negative (downward-entailing) declarative sentences. By examining *individual* participants' *pairs of responses* to both positive and negative plural definite descriptions in gap contexts, we will be able to identify whether they are assigning a homogeneous, existential, or universal interpretation to the plural definite descriptions.

The second main goal of the study is to pursue a characterization of the developmental trajectory to adult homogeneity, by investigating a potential connection with scalar implicatures. We will test Magri's (2014) scalar implicature theory of homogeneity through acquisition, by directly comparing individual children's performance on the two phenomena, using minimally different stimuli. In particular, we will investigate the timeline predictions that the account makes, specifically that we may observe the concurrent emergence of homogeneity and the *some-but-not-all* scalar implicature, or the emergence of the scalar implicature before homogeneity, but crucially not the emergence of homogeneity before the scalar implicature[10].

A final difference we should point out between the previous studies and the present one concerns the tasks presented to the children. The production, act-out, and truth value judgment tasks used in Karmiloff-Smith (1979), Munn et al. (2006), and Caponigro et al. (2012) all involved some degree of reasoning about someone else's desires and actions. On the act-out tasks, children had to satisfy the demands of an issued request; they therefore had to decide how much action they would have to take in order to satisfy the speaker's desires. On the production task, children had to decide how much information to give in order for a third party to successfully carry out an action the way a second party had modeled it. On the TVJT, children had to assess the depicted scenarios against Cookie Monster's likes and dislikes. We make no claims about how adept children are at this kind of reasoning; we will, however, attempt to avoid this extra step entirely, and simplify the task by asking children to judge very simple descriptions of pictures of familiar objects.

4. EXPERIMENT 1

We designed a Truth Value Judgment Task to assess the interpretations that adults and children assign to positive and negative sentences containing plural definite descriptions. Participants' responses to the positive and negative descriptions in GAP contexts would allow us to determine whether they interpreted the definite plural homogeneously, existentially, or universally. To investigate the predictions of the scalar implicature account of homogeneity (Magri, 2014), we also tested participants' interpretation of *some*-sentences in contexts that made the *not-all* implicature false. The direct comparison between homogeneity and scalar implicatures would allow us to assess the potential developmental connection between homogeneity and scalar implicatures.

4.1. Methods

Ethical approval for this study was obtained from the CERES ("Comité d'évaluation éthique des projets de recherche en santé non soumis à CPP") under approval number 2013/46. Written informed consent was obtained from the parents or guardians of all child participants; adult participants were tested through an anonymous web-based survey, and had to click a button to provide informed consent before starting the experiment.

4.1.1. Participants

We tested 24 French-speaking children (13 female) $(4;04,15-5;03,24, M=4;09)$ at two preschools in Paris. Two additional children were excluded because they answered fewer than six of eight control trials correctly (trials in which a sentence

[9] In fact, Caponigro et al.'s proposal would appear to suggest that young children don't initially fall into any of the three categories, or at least not systematically so.

[10] The reader might wonder whether Caponigro et al.'s (2012) *some* conditions could potentially speak to children's ability to compute scalar implicatures. The authors reported the percentage of maximal responses to *some*; notice that a maximal response would be consistent with the literal *some-or-all* meaning of the existential quantifier, but not with the *some-but-not-all* scalar implicature. Given this, it is striking to note that even 4- and 5-year-olds gave very few maximal responses in the *some* condition. This finding cannot be taken as conclusive evidence of calculation of the scalar implicature, however, for a couple of reasons. First, in an act-out task, participants may be driven to take the minimal action to satisfy a request; a less-than-maximal response, which involves less effort, would still be compatible with the literal interpretation of *some*. Second, in any given session, children were presented with requests containing *some* and requests containing *all*. Children could therefore have inferred that the use of *some* should elicit a different response from the use of *all*. Children have indeed been shown to be able to differentiate weak from strong scalar terms; such a finding, however, only establishes sensitivity to relative informativity, and may not necessarily signal the computation of implicatures (for relevant discussion, see Chierchia et al., 2001; Gualmini et al., 2001; Katsos and Bishop, 2011). Finally, we should make it very clear here that Caponigro et al. (2012) do not seek to make any claims about children's scalar implicatures. We simply point out here that although they included an existential quantification condition, we cannot draw on their results to make strong conclusions about children's performance on scalar implicatures.

with a definite description was made uncontroversially true or uncontroversially false). The inclusion criterion of 75% accuracy on controls is fairly standard in truth value judgment task experiments of this kind, and was decided upon prior to testing. We also tested 22 adult native speakers of French, recruited through the online platform FouleFactory, at a total cost of €57.60. All adults passed the controls and were included in the analysis.

4.1.2. Procedure

Children were introduced to a puppet named Raffie the Giraffe, who interacted via webcam. Children were told that Raffie was still very little, and not very good at paying attention. They were then presented with a series of pictures, each containing four objects, and were asked to identify the colors of each of the four objects. The puppet was then asked to say something about the objects, and would utter a test sentence containing a plural definite noun phrase (e.g., *les ballons* "the balloons"), an existentially quantified noun phrase (e.g., *certains ballons* "some balloons"), or a universally quantified noun phrase (e.g., *tous les ballons* "all of the balloons"). Children had to judge the puppet's description and indicate their judgment by stamping on a score sheet, either under a happy face or a sad face.

Children were tested individually away from their classrooms. Responses were videorecorded for subsequent analysis. Children saw two training items involving the description of single, colored objects (i.e., a pink chair and a green piano), followed by 24 test trials presented in one of two pseudorandomized orders, one the reverse of the other (the order of presentation was counterbalanced across participants). The total task took roughly 10 min for children to complete.

Adults were tested on a web-based version of the task; the procedure and the visual stimuli were the same, but the sentences were presented visually (in the form of speech bubbles beside the puppet's picture) rather than orally. Adult participants indicated their responses by clicking on appropriate yes/no buttons.

4.1.3. Materials

As we will describe in more detail below, participants received two training items, six homogeneity targets, eight uncontroversially true/false plural definite description controls, six universal quantification controls, and four scalar implicature targets. The full set of test sentences is provided in the **Appendix**.

Homogeneity targets. Participants heard three positive and three negative *les* "the"-NP sentences such as (12), presented in GAP contexts such as **Figure 3**, in which only two of the four objects in the image were of the color indicated in the test sentence[11].

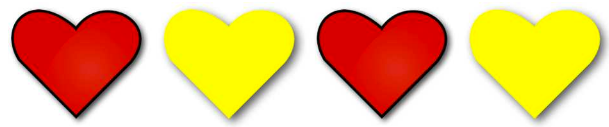


FIGURE 3 | Example of an image presented in the GAP condition. The first and third hearts are red, while the second and fourth hearts are yellow. The corresponding homogeneity target sentence was either *Les coeurs sont rouges* "The hearts are red" or *Les coeurs ne sont pas rouges* "The hearts are not red".

(12) a. Les coeurs sont rouges.
"The hearts are red."
b. Les coeurs ne sont pas rouges.
"The hearts are not red."

If children treated the plural definite description as imposing homogeneity, they were expected to reject both the positive and the negative *the*-sentence, in accordance with the HOMOGENEOUS pattern in **Figure 2**. If children interpreted the definite plural existentially, they were expected to accept the positive *the*-sentence but to reject the negative *the*-sentence. In contrast, if they interpreted it universally, they were expected to reject the positive but to accept the negative sentence. Participants saw three repetitions of the positive definite descriptions and three repetitions of the negative.

Homogeneity controls. In addition to the six homogeneity targets, participants also heard four positive and four negative definite descriptions like (13), presented in contexts that satisfied homogeneity (**Figure 4**); these allowed us to ensure that children understood basic plural definite descriptions, and in particular, could provide *yes*- and *no*-responses appropriately when there were no issues of non-homogeneity. In ALL contexts, where all of the objects shared the same color, the positive control was associated with a *yes*-target, and the negative with a *no*-target. In NONE contexts, where none of the objects had the color indicated in the test sentence, the positive definite description was associated with a *no*-target, and the negative with a *yes*-target.

(13) a. Les parapluies sont rouges.
"The umbrellas are red."
b. Les parapluies ne sont pas rouges.
"The umbrellas are not red."

The targets for these definite control trials were selected dynamically based on children's responses to the target sentences. Every third trial corresponded to a dynamic control, for which the experimenter could select either the *yes*- or the *no*-target. This precaution allowed us to ensure that participants could give both *yes*- and *no*-responses where appropriate, and allowed us to avoid overly long sequences of successive *yes*- and *no*-targets, which otherwise might encourage a *yes*- or *no*-bias, respectively (for previous examples of the use of such dynamic fillers, see Musolino and Lidz, 2006; Conroy et al., 2009; Tieu and Lidz, 2016; Lewis et al., 2017). Any participant who failed to correctly

[11] We varied the objects described and the colors of the objects in order to keep the task engaging for young children. We selected simple objects and colors that preschool-aged children would be familiar with, and ensured that pairs of colors were discernible for any colorblind participants. Children were also asked to verbally identify the objects (*Qu'est-ce que tu vois sur cette image?* "What do you see in this picture?") and the colors of these objects. We did not systematically control for gender but aimed for a rough balance of masculine and feminine nouns across the experiment; of the six critical homogeneity targets, three were feminine (*les voitures*, *les étoiles*, *les balles*) and three were masculine (*les ballons*, *les camions*, *les coeurs*).

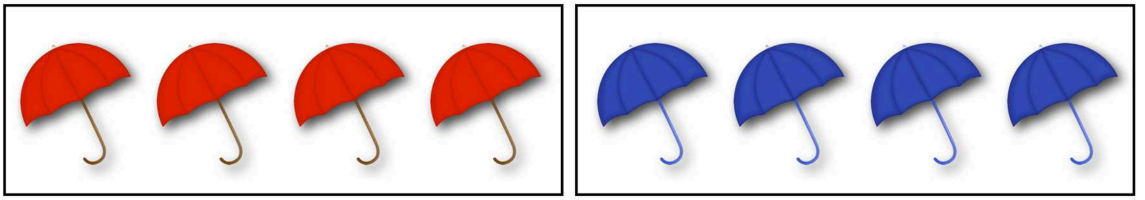


FIGURE 4 | Images corresponding to the plural definite description control condition. When accompanied by the image on the left, in which all four umbrellas are red, the positive and negative descriptions in (13) would be associated with a *yes*- and a *no*-target, respectively. When accompanied by the image on the right, in which all four umbrellas are blue, the positive and negative sentences in (13) would be associated with a *no*- and a *yes*-target, respectively.

answer at least six of the eight definite plural controls was excluded from analysis.

Finally, we included a universal quantification condition, which contained three positive and three negative universally quantified descriptions such as (14), presented in GAP contexts such as **Figure 3**. These would allow us to ensure that children could assign an adult-like, negation-preserving meaning to universally quantified sentences, and would provide a point of comparison for the plural definite descriptions.

(14) a. Tous les coeurs sont rouges.
"All the hearts are red."
b. Pas tous les coeurs sont rouges.
"Not all the hearts are red."[12]

Scalar implicature targets. To assess Magri's (2014) scalar implicature-based account of homogeneity, we also administered a scalar implicature test. Participants received four scalar implicature trials, which involved existentially quantified *certains* "some"-sentences, presented in contexts where all four objects displayed were of the mentioned color (**Figure 5**). If participants computed the *some-but-not-all* implicature, they were expected to reject the test sentences. If they accessed only the literal plain existential meaning of the sentences, however, they were expected to accept the descriptions. This condition would allow us to directly compare participants' performance on homogeneity and scalar implicatures.

Summary of the materials. In all, participants received two training items, six homogeneity targets, eight uncontroversially true/false plural definite description controls, six universal quantification controls, and four scalar implicature targets. The full set of test sentences is provided in the **Appendix**.

FIGURE 5 | Example of an image presented in the scalar implicature target condition. All four of the tents are orange, while the corresponding test sentence was *Certaines tentes sont oranges* "Some tents are orange."

4.2. Results

4.2.1. Plural Definite Descriptions

Figure 6 displays the percentage of *yes*-responses for the homogeneity targets, in between the true and false definite description controls. While children were adult-like with respect to the definite description controls, the two groups differed in their treatment of the definite plural in GAP contexts. Unlike the adults, the children showed some acceptance of the positive definite descriptions in gap contexts; a mixed effects logistic regression model of responses as predicted by polarity revealed that they accepted the positive targets significantly more than they did the negative targets ($p < 0.001$) (lme4 package for R, R Core Team, 2016, Bates et al., 2015).

That children behaved differently from adults indicates that the child participants as a group were non-adult-like in their interpretation of the definite plural; but we wished to explore further *how* they might be interpreting the plural definite descriptions. Previous studies have hinted at existential, universal, and homogeneous possibilities, but these studies were inconclusive in this respect due to the absence of negative definite description targets. To further explore the possible interpretive preferences, we categorized participants according to their responses to both the positive and negative homogeneity targets. A participant was categorized as exhibiting the HOMOGENEOUS response pattern if they rejected at least two of three positive homogeneity targets and at least two of three

[12] Some native speakers of French may not find the negative (14b) to be an entirely natural formulation. We chose to place the negation before the universal quantifier for three reasons. First, a natural alternative would have been to float the universal quantifier, as in *Les coeurs ne sont pas tous rouges* "The hearts are not all red." But we chose to avoid any potential issues related to children's mastery of floating quantification. Second, if we kept the universal quantifier in its unfloated position, as in *Tous les coeurs (ne) sont pas rouges* "All of the hearts are not red," there was a question of whether children might be sensitive to the presence or absence of the *ne*. If, for whatever reason, participants failed to perceive the *ne*, for example, this could have invited an interpretation where the negation was phrasal rather than sentential, e.g., *Les coeurs sont [pas-rouges]* "The hearts are [not-red]." Finally, the version where *pas* precedes *tous* gives rise to the rather strong indirect scalar implicature that some of the hearts are red, which would serve to further emphasize the absence of homogeneity depicted in the test image.

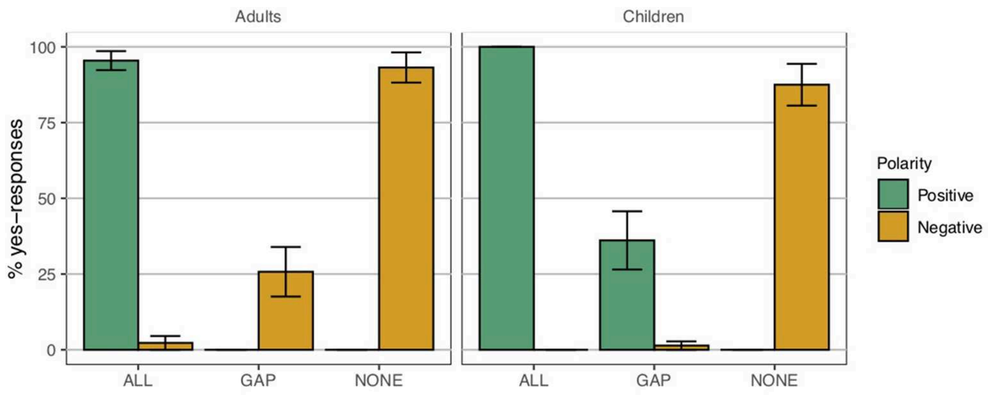


FIGURE 6 | Percentage of *yes*-responses to the plural definite description targets and controls. Error bars correspond to the standard error of the within-participant means. Homogeneity targets corresponded to plural definite descriptions of GAP contexts. Clearly true controls corresponded to positive plural definite descriptions of ALL contexts and negative plural definite descriptions of NONE contexts. Clearly false controls corresponded to positive plural definite descriptions of NONE contexts and negative plural definite descriptions of ALL contexts.

TABLE 1 | Distribution of participants according to their performance on homogeneity and scalar implicature targets.

	Adults		Children	
	− Implicature	+ Implicature	− Implicature	+ Implicature
Homogeneous	5	10	6	10
Existential	0	0	7	1
Universal	5	1	0	0

negative homogeneity targets. A participant was categorized as exhibiting the EXISTENTIAL response pattern if they accepted at least two of three positive homogeneity targets, and rejected at least two of three negative homogeneity targets. Finally, a participant was categorized as displaying the UNIVERSAL response pattern if they rejected at least two of three positive homogeneity targets, and accepted at least two of three negative homogeneity targets[13].

Table 1 displays the distribution of participants in the different response categories, based on their performance on the homogeneity and scalar implicature targets. Some readers would prefer an alternative analysis that does not bin participants into categories; we include this discussion here as an exploration of the possible interpretive profiles. As it turns out, our participants aligned rather strikingly into a subset of the possible categories. Let us first focus on the homogeneity targets. Sixteen of the 22 adult participants displayed the HOMOGENEOUS pattern of responses, rejecting both positive and negative definite descriptions in GAP contexts, while six adult participants displayed the UNIVERSAL response pattern, accepting the negative targets but rejecting the positive targets. Children treated the homogeneity targets differently from the adult group: sixteen of the 24 children displayed the HOMOGENEOUS pattern of responses, while the remaining eight children displayed the EXISTENTIAL response pattern ($\chi^2(2, N = 46) = 13.94, p < 0.001$). No adult displayed the EXISTENTIAL response pattern and no child displayed the UNIVERSAL response pattern.

We also elicited follow-up justifications following children's responses. The explanations that children provided indicate that they were generally consistent in their responses to the target conditions. Children consistently rejected the negative plural definite descriptions in GAP contexts, justifying their answers by pointing out the objects that had the color mentioned by the puppet, as in (15).

(15) Justifications for rejecting negative homogeneity targets

a. *Les camions ne sont pas bleus.* "The trucks are not blue"
CHI: (Non) parce qu'il y en a des bleus
"(No) because there are blue ones" (C03-A, age 4;09,20)

b. *Les camions ne sont pas bleus.* "The trucks are not blue"
CHI: (Non) parce que les camions ils sont bleus et jaunes
"(No) because the trucks are blue and yellow" (C05-A, age 4;09,21)

c. *Les balles ne sont pas rouges.* "The balls are not red"
CHI: Pas vrai. Il y en a qui sont rouges
"Not true. There are some that are red" (C07-B, age 4;11,19)

The *yes*-responses observed in the positive definite GAP condition were primarily elicited from eight children who consistently

[13] We chose to have three repetitions of each of the positive and negative homogeneity targets so as to keep the overall length of the experiment manageable for children; but a future study might ideally include more target trials, to ensure that a participant's categorization under this scheme truly reflects their interpretation of the definite description. In this respect, however, it is worth noting that our child participants were remarkably consistent in their responses to the targets, with only two of the children giving non-uniform responses (to the positive targets).

accepted in this condition. These children justified their *yes*-responses by pointing out the objects that were of the color mentioned by the puppet, as in (16).

(16) Justifications for accepting positive homogeneity targets

a. *Les ballons sont rouges.* "The balloons are red"
CHI: (Oui) elle a dit qu'ils sont rouges
"(Yes) she said they're red" (C02-B, age 4;04,15)

b. *Les ballons sont rouges.* "The balloons are red"
CHI: (Oui) parce qu'il y en a deux rouges
"(Yes) because there are two red ones" (C03-A, age 4;09,20)

c. *Les voitures sont bleues.* "The cars are blue"
CHI: (Oui) elle a raison, elle a dit les voitures elles sont bleues
"(Yes) she's right, she said the cars are blue" (C09-A, age 4;05,09)

The HOMOGENEOUS children who rejected the positive homogeneity targets justified their responses by drawing attention to the objects that were of the color not mentioned by the puppet, as in (17).

(17) Justifications for rejecting positive homogeneity targets[14]

a. *Les ballons sont rouges.* "The balloons are red"
CHI: (Non) parce qu'ils sont rouges et bleus
"(No) because they are red and blue" (C05-A, age 4;09,21)

b. *Les étoiles sont jaunes.* "The stars are yellow"
CHI: (Non) parce qu'elle a oublié les rouges !
"(No) because she forgot the red ones" (C07-B, age 4;11,19)

c. *Les voitures sont bleues.* "The cars are blue"
CHI: (Non) c'est pas tout bleu
"(No) it's not all blue" (C12-B, age 4;07,14)

4.2.2. Scalar Implicatures

Children's performance in the scalar implicature condition was comparable with that of the adult participants: children rejected the existentially quantified descriptions of ALL contexts 46% of the time, while adults did so 50% of the time. The distribution of adult and child participants according to their performance on homogeneity and scalar implicature targets is summarized in **Table 1**. An examination of individual children's responses in this condition revealed two groups of children: those who consistently failed to compute the implicature, accepting on at least three of four implicature trials, and those who consistently computed the implicature, rejecting on at least three of four implicature trials. Eleven of the 24 children consistently computed implicatures, providing justifications consistent with the strengthened meaning of the sentences:

(18) Justifications consistent with calculation of scalar implicature

a. *Certains chapeaux sont roses.* "Some hats are pink"
CHI: (Non) tous les chapeaux sont roses
"(No) all of the hats are pink" (C10-B, age 4;10,12)

b. *Certains chapeaux sont roses.* "Some hats are pink"
CHI: (Non) parce qu'elle a dit certains [...] j'aurais dit qu'ils sont tous roses
"(No) because she said some [...] I would have said they're all pink" (C11-A, age 5;00,05)

c. *Certaines tentes sont oranges.* "Some tents are orange"
CHI: Oh non, parce qu'elles sont toutes oranges
"Oh no, because they're all orange" (C13-B, age 4;09,16)

In all, 13 of the 24 children failed to compute scalar implicatures, accepting on at least three of the four scalar implicature trials. Seven of these children were among the eight children who displayed the EXISTENTIAL response pattern to the homogeneity targets, accepting the positive homogeneity targets and rejecting the negative ones[15]. The other six children who failed to compute implicatures were a subset of the 16 children who displayed the HOMOGENEOUS response pattern.

4.2.3. Non-randomness of Groupings

One potential concern about the groupings reported above is that some children, having not yet acquired the relevant construction, simply answered randomly (that is, at chance) on the homogeneity targets or the implicatures targets, or both, and therefore our diagnosis of a group of children with homogeneity but no implicatures may be spurious.

Recall that there were three items per condition and participants were categorized by their majority response. Based on the two homogeneity target conditions (THE-SOME-POS and THE-SOME-NEG), every participant is thus guaranteed to fall into one of four possible groups. The fourth group was not mentioned in the preceding discussion because it turns out to be empty and is the least plausible from a theoretical point of view: it would correspond to interpreting the definite description as an existential that takes scope above negation. Now, given that two of the six possible groups are empty, it would be rather surprising if all six of the homogeneous/−implicature children ended up in this group by giving random responses, without any child ending up in one of the two empty groups (where they could have landed just as well by answering randomly).

To put a number on it, assume the following. Take children's answers on implicature targets to be non-random. This means we can exclude the +implicature children from consideration, since they cannot, in virtue of randomness of their responses to homogeneity targets, end up in the homogeneous/−implicature group. Now assume that of the 13

[14]Some of these justifications could be consistent with rejections not for non-homogeneity but rather for a failure to completely describe all of the objects in the picture. That is, the puppet only accurately described half of the objects. It is difficult to tease apart the two kinds of rejections here. Experiment 2 will include a control that allows us to address this potential concern.

[15]Unlike the child participants, no adult displayed the EXISTENTIAL response pattern. Not all adults computed scalar implicatures, however. We will return to this point.

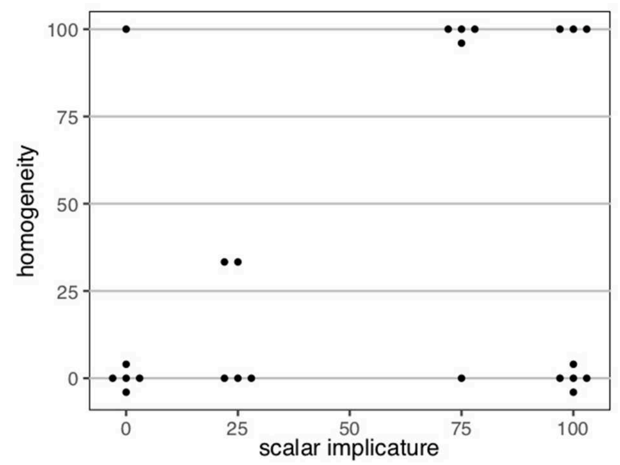


FIGURE 7 | Within-subject mean responses on affirmative homogeneity and implicature items for children in Experiment 1 (recall that rejection (0) on the task corresponded to +homogeneous and +implicature responses).

−implicature children, a certain number n answered randomly on homogeneity targets. Since the hypothesis is that the whole homogeneous/−implicature group is spurious, the value of n has to be at least 6. Now consider the probability, as a function of n, that the results would be at least as extreme as they actually are, in the following sense: at least six children are categorized as homogeneous/−implicature, and the other children are categorized as existential/−implicature, while the other two possible groups are empty. (The remaining $13 - n$ non-random responders fall in the existential/−implicature group in any case.) We find that for all values of n in $[6, 13]$, with the exception of $n = 8$, $p < 0.0005$ (for $n = 8$, $p < 0.001$). Alternatively, assume that the random responders answered randomly on both homogeneity and implicature targets. We consider a result to be at least as extreme as ours if the following is the case: at least six children are in the homogeneous/−implicature group, no children are in the universal or the wide-scope existential group, and at most one child is in the existential/+implicature group. Then for any value of n in $[6, 24]$, $p < 0.0001$.

To see the point in a more visual form, consider **Figure 7**, which shows individual children's mean responses to the positive homogeneity and implicature targets, with each data point corresponding to an individual child. The four corners correspond to groups: existential/+implicature in the upper-left, existential/−implicature in the upper-right, homogeneous/+implicature in the lower-left, and homogeneous/−implicature in the lower-right corner. Observe that children do, indeed, cluster into the corners nicely and the center of the plane is empty, indicating that children's responses are systematic and not random, legitimizing the binning into groups.

We may thus safely conclude that our finding is not an artifact created by children simply giving random responses to homogeneity targets coupled with a categorization rule that is based on the majority response in an odd number of trials.

TABLE 2 | Predicted responses to each condition ([determiner]-[context]-[polarity]) for each of the six possible groups, defined by the reading for the definite (EXIstential, HOMogeneous, UNIversal) and the presence or absence of implicatures.

	EXI		HOM		UNI	
	+SI	−SI	+SI	−SI	+SI	−SI
THE-ALL-NEG	0	0	0	0	0	0
ALL-GAP-POS	0	0	0	0	0	0
ALL-GAP-NEG	1	1	1	1	1	1
THE-NONE-POS	0	0	0	0	0	0
SOME-ALL-POS	0	1	0	1	0	1
THE-NONE-NEG	1	1	1	1	1	1
THE-GAP-POS	1	1	0	0	1	1
THE-GAP-NEG	0	0	0	0	1	1
THE-ALL-POS	1	1	1	1	1	1

4.2.4. Improved Group Assignment

The purpose of this section is to provide a more solid underpinning for the descriptive categorization of participants we gave above. While in the previous section, we established that the observed group assignment is highly unlikely to be the result of purely random responses, our child data are clearly quite noisy, which a simple categorization based on majority response does not take into account. This issue will become especially pressing in Experiment 2, where the number of possible groups is much larger. We thus performed a categorization of participants on the basis of a statistical model of the responses.

The task is to assign a *group* to every participant, where there are 6 possible groups determining (i) whether or not the participant derives implicatures and (ii) what reading this participant assigns to definite descriptions (HOMogeneous, EXIstential, UNIversal). A group thus determines a theoretical response to each condition, as described in **Table 2**. We fitted logit models of the data (including both target and control conditions), with fixed intercept and slope and a subject-dependent group parameter as a predictor variable, varying by subject[16]. The probability that a participant belongs to a given group is then given by the posterior probability of that value of the group parameter for that participant[17].

For children, the model fitted with all six possible levels for the group predictor indicated no mentionable posterior probability of a universal reading of the definite plural for any child.

[16]The addition of varying intercepts and/or slopes in the models led to convergence problems and was therefore eschewed.

[17]Technically, the model had the form:

$$Y_{s,i} \sim \text{bernoulli}(\text{logit}^{-1}(\pi_{s,i})), \text{ with } \pi_{s,i} = \alpha + \beta X_{\gamma(s)i} \text{ and } X_{gi} \text{ following } \textbf{Table 2}.$$

Bayesian models were fitted separately for children and adults using JAGS through the `rjags` package (Plummer, 2003). The prior for the intercept parameter α was set to a normal distribution with mean 0 and precision 0.001, while the prior for the slope parameter β was the non-negative half of the same distribution (since a participant cannot plausibly be more likely to judge a sentence true when it is, in fact, false). The prior for $\gamma(s)$, the group parameter of each participant s, was uniform. 5,000 samples were drawn from each of 4 chains after 5,000 burn-in iterations.

TABLE 3 | Count of group assignments ($\arg\max_g p(\gamma(s) = g|Y)$) for children in Experiment 1.

	−SI	+SI
HOM	6	10
EXI	7	1

TABLE 4 | Estimated log pointwise predictive likelihood (elpd) and its standard error with different available group assignments for children.

Possible groups	elpd	(se)
All	−130.30	(15.86)
All but EXI/+SI	−137.97	(15.98)
All but HOM/−SI	−181.09	(15.90)
All but EXI/+SI and HOM/−SI	−189.79	(15.90)

"All" groups were EXI/±SI and HOM/±SI.

The two corresponding groups were thus subsequently dropped and the analysis was re-run with only four possible values for the group parameter. Children were assigned to groups quite unambiguously: the posterior probability of the group with the highest posterior probability ($\max_g p(\gamma(s) = g|Y)$) was > 0.92 for all children and > 0.99 for all except two. The result, shown in **Table 3**, replicates exactly our descriptive categorization[18].

One might want to evaluate more directly whether the HOM/−SI and the EXI/+SI groups can be assumed to be populated. To do so, we compared models which made these groups a possibility with models which did not, using a leave-one-out cross-validation as recommended by Vehtari et al. (2017)[19]. **Table 4** summarizes the obtained estimated log pointwise predictive likelihoods. We see that models with the HOM/−SI group perform much better than those that do not include it (e.g., with all other groups included, $\Delta_{\text{elpd}} = 50.79$ with $se = 9.87$), showing that this group is indeed populated. In comparison, models including the EXI/+SI group outperform their counterparts without it by only a small margin (e.g., $\Delta_{\text{elpd}} = 7.67$ with $se = 5.08$).

The model for adults was also first fitted with all six groups, followed by dropping the possibility of an existential reading since the model was found not to make use of it. Group assignment was again quite unambiguous[20]. The results are shown in **Table 5**; they are qualitatively comparable to our descriptive categorization from the previous section.

4.3. Discussion

Let us first consider the results from the adult participants. Adult subjects were about equally split between those who did and those

TABLE 5 | Count of group assignments ($\arg\max_g p(\gamma(s) = g|Y)$) for adults in Experiment 1.

	−SI	+SI
HOM	8	11
UNI	3	0

who did not derive scalar implicatures. This is not surprising given that implicatures are often said not to be obligatory[21] and participants have previously been found to vary in the rate of implicature-based responses in such tasks (see e.g., Noveck and Posada, 2003).

As for the definite descriptions, the overwhelming majority of adults interpreted them homogeneously and none treated them as existential, as we would expect. In addition, a small number of participants treated the plural definite description like a (low-scope) universal; that is to say, in GAP situations they judged affirmative THE-sentences false, but negated ones true.

One can think of various possible explanations for this. One is that the definite description is really a universal for all speakers, but some chose the wide-scope and some chose the low-scope reading in a scopally ambiguous case, such as that of sentential negation. Since, however, there are independent arguments for why homogeneity is not simply universally interpreted definite plurals taking wide scope[22], this has little plausibility.

Alternatively, these particular participants might just have a different understanding of the definite from the majority, namely a universal as opposed to a homogeneous one. This hypothesis would be quite testable precisely on the basis of the arguments for a distinction between homogeneity and wide-scope universals, since these participants would be predicted to behave distinctly on such cases. However, we do not pursue this question further here.

Finally, these participants might be employing a different response strategy: instead of first computing the truth value of the sentence in a trivalent setting and then mapping these three truth values to two truth values to generate their response, they might, following the intuition that negation should invert the truth value, first compute their response for the positive sentence and then simply reverse it to obtain the response for the negated sentence[23].

Turning to the children's responses, recall that the first goal of the experiment was to resolve the uncertainty surrounding the interpretations that young children assign to plural definite descriptions; previous studies had hinted at existential, universal, and homogeneous possibilities, but these hints were inconclusive due to the absence of the negative counterparts. The results of Experiment 1 revealed two groups of children, based on

[18]The group assignments remained the same when the model was fitted only on the items directly relevant to homogeneity and implicatures, i.e., the items in the THE-GAP-POS, THE-GAP-NEG, and SOME-ALL-POS conditions.

[19]Deviating from Vehtari et al. (2017), we did not approximate the cross-validation by importance sampling, as it was questionable that our data set would meet the prerequisites for this procedure.

[20]$\min_s(\arg\max_g p(\gamma(s) = g|Y))$, i.e., the minimal probability with which any subject was assigned its group was 0.82. The mean was 0.96.

[21]At least not regardless of context, and participants may differ in what kind of context they assume to obtain in the experimental situation.

[22]The relevant evidence comes from definite plurals embedded in non-monotonic contexts, as well as definite plurals with a pronoun bound by a higher negative quantifier (cf. Magri, 2014 and Križ and Chemla, 2015).

[23]This is equivalent to interpreting negation as weak negation ($\sim p$ is true if p is not true) instead of strong negation ($\neg p$ is true if p is false) in a trivalent logic.

responses to both positive and negative definite descriptions: one group interpreted the definite descriptions existentially (scoping under negation), while the other interpreted them homogeneously. We had initially reasoned that a universal interpretation would be plausible on the basis of considerations of the input. If a child were to hear positive plural definite descriptions exclusively in scenarios that satisfied homogeneity, for example, that could be a strong reason to posit a universal meaning for the definite plural. The fact that no child displayed the UNIVERSAL pattern of response, however, suggests this is not the case. Instead, children might be led to posit an existential meaning for the definite plural, on the basis of its behavior under negation, and the occasional non-maximal reading of the definite plural (for discussion of non-maximal readings, see Brisson, 1998; Lasersohn, 1999; Malamud, 2012; Schwarz, 2013; Križ, 2015a).

Note another important finding of Experiment 1. While non-maximal responses from children could be argued to arise from non-adult-like domain restriction, the inclusion of negative targets in our experiment allows us to rule out such an explanation for their seemingly existential readings of plural definite descriptions. If children (in our experiment as well as in the previous experiments we've discussed) were to accept the positive plural definite descriptions in a gap scenario because they restricted the domain to the individuals that did indeed satisfy the predicate, one would expect them to be able to accept the negative homogeneity targets using an analogous strategy of restricting the reference to those individuals who do not satisfy the predicate. In essence, such 'wildly domain-restricting' children would interpret the positive and negative homogeneity targets as in (19) and (20), respectively.

(19) The hearts are red.
$\rightsquigarrow$ The hearts that are red are red.

(20) The hearts are not red.
$\rightsquigarrow$ The hearts that are not red are not red.

The fact that the children we tested, in particular those who accepted the positive homogeneity targets, never accepted the negative targets, suggests that acceptance of homogeneity violations cannot be due to non-adult-like domain restriction.

The second goal of the experiment was to investigate the predictions of the scalar implicature account of homogeneity. On this account, the definite plural has a literal existential meaning, which is then strengthened to a universal meaning through an implicature. The finding of an EXISTENTIAL subgroup of children, who moreover lacked scalar implicatures, is consistent with and expected on the implicature account of homogeneity. Unable to derive the homogeneous meaning through implicature, these children start out with judgments based on the literal, existential meaning of the definite plural.

The implicature account also makes the further prediction, however, that homogeneity should not be observed in the absence of scalar implicatures. This prediction comes in two parts. First, children who have not yet acquired scalar implicatures should be unable to obtain homogeneous readings for plural definite descriptions. Second, the scalar implicature from *some* to *not all* should not occur at a lower rate than homogeneous readings because this implicature is actually a subcomputation of the homogeneity implicature in Magri's theory. If anything, homogeneity should occur at a lower rate than the regular scalar implicature.

Even among our adult participants, roughly half were categorized as not computing implicatures. This means that we cannot conclude that the children who are categorized as not deriving implicatures have indeed not yet acquired them, since it is also possible that they simply refrain from computing implicatures for the same reason that some of the adults do. Consequently, our data do not speak to the first prediction of the implicature theory. The second prediction, however, is clearly falsified for both children and adults: in both groups, the failure to derive scalar implicatures is more prevalent than non-homogeneous interpretations of plural definite descriptions[24]. Most strikingly, there was no group of participants who systematically derived scalar implicatures and at the same time failed to access homogeneous readings of definite plurals. This suggests that there is, in fact, an alternative way of obtaining homogeneous readings that does not rely on scalar implicatures, and that this alternative way of generating homogeneity is already acquired by the time children are robustly computing scalar implicatures.

A remaining worry is that our diagnosis of universal and homogeneous readings might be confounded by the scope of negation. The present analysis is predicated on the assumption that the definite plural, whatever its meaning, takes low scope under sentential negation. However, in order to keep the sentences and visual display simple, we had the definite plural in the subject position of intransitive sentences, which means that its surface scope was actually above sentential negation. If children interpreted the definite plural as a universal in surface scope position, i.e., with wide scope over negation, then that would give rise to the same responses as a homogeneous meaning in our binary judgment paradigm: both affirmative and negative sentences with definite plurals would be judged false (i.e., non-true) in gap situations. As there is not much of a difference in either the mean or youngest age of participants in the existential vs. the homogeneous group (mean 4.72 years and minimum 4.37 years vs. mean 4.73 years and minimum 4.42 years), it is possible that some children start out with a (low-scope) existential reading and others start out with a (wide-scope) universal reading for the definite plural. Experiment 2 is an attempt to control for this possibility.

Note that a wide-scope universal is not under discussion as a possible reading of the definite plural for adults. This reduces the plausibility of the above worry for children, and makes it entirely inapplicable to our argument against the implicature theory on the basis of the adult data.

[24]The nature of the seemingly universal responses from some adults is irrelevant to this argument.

5. EXPERIMENT 2

The goal of Experiment 2 was to obtain a more fine-grained picture in which truly homogeneous readings would be distinguished from wide-scope universals. In order to do this, what we require is a way to distinguish merely non-true sentences from those that are *bona fide* false, to which end a ternary response paradigm has been employed for adults by Križ and Chemla (2015). A ternary response paradigm has also been used with children in an investigation of scalar implicatures. Katsos and Bishop (2011) report that when given the choice between a minimal, an intermediate, and a maximal reward option, 5-year-old children are adult-like in consistently choosing to give the puppet the intermediate reward for a literally true utterance with a false implicature. We were thus hopeful that a similar implementation of the ternary response paradigm would allow us to shed further light on the interpretations that children assign to plural definite descriptions.

FIGURE 8 | Materials used in the ternary judgment task. Clearly false targets were meant to elicit the minimal reward of one strawberry. Clearly true targets were meant to elicit the maximal reward of three strawberries. Based on the results reported in Katsos and Bishop (2011), the intermediate reward of two strawberries was meant to correspond to underinformative targets.

5.1. Methods

Ethical approval for this study was obtained from the CERES ("Comité d'évaluation éthique des projets de recherche en santé non soumis à CPP") under approval number 2013/46. Written informed consent was obtained from the parents or guardians of all child participants; adult participants were tested through an anonymous web-based survey, and had to click a button to provide informed consent before starting the experiment.

5.1.1. Participants

We tested 24 French-speaking children (10 female) (4; 07, 04 − 6; 04, 13, M = 5; 03) at a preschool in Paris. Three additional children did not finish the task, and another two were excluded from analysis because they answered fewer than six of eight control trials correctly (using the same control trials as in Experiment 1, in which a sentence with a definite description was made uncontroversially true or false). We also tested 25 adult native speakers of French, recruited through the online platform FouleFactory, at a total cost of €38.30. All adult participants passed the controls and were included in the analysis.

5.1.2. Procedure

Children were introduced to Boba the puppet, who interacted via webcam. Children were told that Boba was still very little, and not very good at paying attention. Children were then presented with a series of pictures on a laptop computer, each containing four objects, just as in Experiment 1. They were asked to identify the colors of each of the four objects. The puppet was then asked to say something about the objects, and would utter a test sentence containing a plural definite description (e.g., *les ballons* "the balloons"), an existentially quantified noun phrase (e.g., *certains ballons* "some balloons"), or a universally quantified noun phrase (e.g., *tous les ballons* "all the balloons"). Children had to decide whether the puppet's description was worth a reward of one, two, or three strawberries. Children indicated their choices by choosing cards with the appropriate number of strawberries on them and placing them in a box in front of the laptop (**Figure 8**).

Although children in this age range have been reported to engage quite naturally with these kinds of graded reward scales (Katsos and Bishop, 2011), some time at the beginning of the experiment was devoted to making sure each child understood how to use the scale. The instructions for each child included an explanation of how to use the graded reward scale, and the child was encouraged to explain back to the experimenter what each reward meant, to make sure they had understood. Only once the child showed a solid understanding of the three possible rewards did the task begin. The instructions are provided in the **Appendix**, in both French and English.

Children were tested individually away from their classrooms. Responses were videorecorded for subsequent analysis. Children saw two training items containing single objects (e.g., a pink chair), followed by 26 test trials presented in one of two pseudorandomized orders (the reverse of each other). The total task took roughly 10–15 min for each child to complete.

Adults were tested on a web-based version of the task; sentences were presented visually in the form of speech bubbles, and adults indicated their responses by clicking on appropriate buttons depicting the three reward options.

5.1.3. Materials

The materials used in Experiment 2 took essentially the same form as those in Experiment 1, but some additional control conditions were required because of the nature of the judgment task. Recall that the primary goal of this experiment was to tease apart homogeneous readings of definite descriptions from wide-scope universals by giving participants an intermediate response option that could be used to indicate a homogeneity violation.

This is complicated by the fact that an intermediate response to a homogeneity target may conceivably arise for any of the following reasons:

(21) Possible sources of an intermediate reward for a positive homogeneity target, e.g., *The hearts are red* in a GAP context
 a. The child interpreted the definite description homogeneously.
 b. The child interpreted the definite description existentially (*Some of the hearts are red*), but didn't want to give the maximal reward because the sentence was an incomplete description of the image.
 c. The child interpreted the definite description universally (*Every heart is red*), but didn't want to give the minimal reward because the sentence was a true description of at least part of the image.

Likewise, an intermediate reward for a negative homogeneity target would ideally reflect a child's sensitivity to the violation of homogeneity. But it could arise for any of the reasons in (22).

(22) Possible sources of an intermediate reward for a negative homogeneity target, e.g., *The hearts are not red* in a GAP context
 a. The child interpreted the definite description homogeneously.
 b. The child interpreted the definite description existentially (*Some of the hearts are not red*), but didn't want to give the maximal reward because the sentence was an incomplete description of the image.
 c. The child interpreted the definite description universally (*Every heart is not red*), but didn't want to give the maximal reward because the sentence was true on only one of the two possible scopal construals.

To address these potential confounds, we included three kinds of controls in this experiment: *incomplete description* existential controls, *partial truth* universal controls, and *scope ambiguity* universal controls. If a child did not give intermediate responses in these conditions, then we could exclude these three confounds as potential explanations for intermediate responses to the homogeneity targets. The specific sentence types used to control for these three confounds will be described in the appropriate sections below, alongside the corresponding target sentence types.

Plural definite descriptions. Experiment 2 included positive and negative sentences containing plural definite descriptions, as in (23). They were combined with different types of situations (pictures) to form homogeneity targets, as well as clearly true and clearly false controls.

(23) a. Les coeurs sont rouges.
 "The hearts are red."

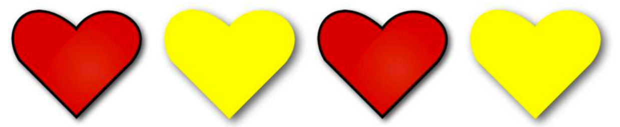


FIGURE 9 | Image corresponding to a GAP context. The first and third hearts are red, while the second and fourth hearts are yellow. If used on a homogeneity target trial, this image would accompany either the positive *Les coeurs sont rouges* "The hearts are red" or the negative *Les coeurs ne sont pas rouges* "The hearts are not red." If used on an incomplete description control trial, this image would accompany the sentence *Certains coeurs sont rouges* "Some hearts are red." If associated with a partial truth control, this image would accompany the positive *Tous les coeurs sont rouges* "All the hearts are red." Finally, if associated with a scope ambiguity control, this image would accompany the negative *Tous les coeurs ne sont pas rouges* "Not all the hearts are red."

 b. Les coeurs ne sont pas rouges.
 "The hearts are not red."

Participants received three positive and three negative homogeneity target trials. On these target trials, they had to judge positive and negative *les* "the"-NP sentences such as (23), presented in GAP contexts in which only two of the four objects in the image were of the color indicated in the test sentence (**Figure 9**).

Participants also received four clearly true or clearly false positive definite description controls, and four clearly true or clearly false negative definite description controls. On these control trials, participants heard sentences containing plural definite descriptions just like (23), but presented in contexts that satisfied homogeneity, i.e., where all four objects displayed were of the same color (**Figure 10**). In ALL contexts, where all of the objects shared the color indicated in the test sentence, the positive control (23a) was associated with a maximal reward target, and the negative control (23b) with a minimal reward target. In NONE contexts where none of the objects had the color indicated in the test sentence, the positive definite description (23a) was associated with a minimal reward target, and the negative (23b) with a maximal reward target.

Whether a definite plural control sentence was accompanied by an ALL or a NONE picture was determined dynamically, on the basis of children's responses to the target trials[25]. This allowed us to avoid eliciting overly long sequences of the same response (for example, a string of successive intermediate rewards), which otherwise could have encouraged a biased response strategy. These controls also allowed us to ascertain that children understood definite descriptions, and in particular could provide minimal and maximal reward judgments appropriately when there were no issues of non-homogeneity. Any participant who failed to correctly answer at least six of the eight definite plural controls was excluded from analysis.

Existential quantification conditions. Experiment 2 also contained positive existentially quantified sentences such

[25]For adult participants, half of the controls involved the ALL context and half involved the NONE context.

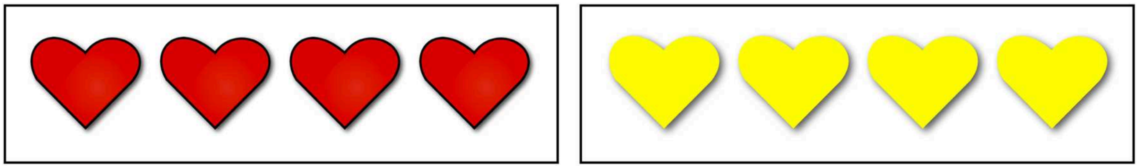


FIGURE 10 | Images corresponding to the clearly true and clearly false definite plural controls. When accompanied by the ALL context image on the left, in which all four hearts are red, the positive and negative descriptions in (23) would be associated with a maximal reward target and a minimal reward target, respectively. When accompanied by the NONE context image on the right, in which all four hearts are yellow, the positive and negative sentences in (23) would be associated with a minimal reward target and a maximal reward target, respectively.

as (24). They were combined with two types of situations (pictures) to form scalar implicature targets and incomplete description controls.

(24) Certains coeurs sont rouges.
"Some hearts are red."

On scalar implicature trials, participants heard such sentences in contexts where all four objects displayed were of the mentioned color. Each participant received three such trials. As with the homogeneity targets, we expected that if participants computed the scalar implicatures, they would opt to give either minimal or intermediate rewards, but not maximal rewards. This is because although the sentences are true on their literal meaning, the context falsifies the associated scalar implicatures. Previous work by Katsos and Bishop (2011) suggests that children are likely to give intermediate rewards for such cases of underinformative descriptions.

On the *incomplete description* controls, participants heard existentially quantified sentences as descriptions of GAP contexts. For example, they would hear a sentence like (24), accompanying the image in **Figure 9**[26]. These sentences are uncontroversially true in such contexts, but they do not offer a complete description of the situation and here quite visibly so: a color present in the picture is not at all mentioned in the sentence. So if a participant gives an intermediate reward on these control trials, we may suspect that other intermediate rewards they might give for homogeneity targets could also be due to incomplete description effects. Each participant received three repetitions of this control.

Universal quantification conditions. Finally, Experiment 2 also contained positive and negative universally quantified sentences, as in (25). These were combined with GAP contexts to form partial truth controls and scope ambiguity controls.

(25) a. Tous les coeurs sont rouges.
"All the hearts are red."
b. Tous les coeurs ne sont pas rouges.
"All the hearts are not red."
Intended interpretation: "Not all the hearts are red."[27]

On partial truth controls, positive universally quantified sentences such as (25a) were presented in GAP contexts like **Figure 9**, in which only two of the four objects were of the color indicated in the test sentences. Each participant received three such trials. These sentences were uncontroversially false in GAP contexts, so if a participant gave an intermediate reward rather than a minimal reward, we could reasonably infer that they had a bias for rewarding the puppet for having given a truthful description of at least part of the picture. This would then give us reason to suspect that any intermediate responses the participant may have given on the homogeneity targets could also have arisen from these partial truth effects.

On scope ambiguity controls, a negative universally quantified sentence such as (25b) was presented in a GAP context like **Figure 9**. Each participant received three such trials. On the intended interpretation, the negative sentences were true in GAP contexts. On the other construal, on which the universal scopes above negation, the sentences were false. If a child gave an intermediate reward rather than a maximal reward, this could reflect a dispreference against sentences that had at least one false reading. In other words, the puppet would receive a reward for saying something that had a true reading, but would not receive the maximal reward because the utterance was not *unambiguously* true. This would then give us reason to suspect that any intermediate responses the participant may have given on the negative homogeneity targets could also have been given on the grounds of a scope ambiguity between a (universally or existentially interpreted) definite description and negation[28].

Summary of the materials. In all, participants received two training items, six homogeneity targets, eight uncontroversially true/false plural definite description controls, three scalar implicature targets, three incomplete description controls, three partial truth controls, and three scope ambiguity controls. The full set of test sentences is provided in the **Appendix**.

5.2. Results

5.2.1. Existential Quantification Conditions

Figure 11 displays the percentages of the reward types given in the existential quantification conditions. In response to the

[26]This control also allowed us to ensure that the children could access adult-like interpretations of existentially quantified sentences.

[27]We provide the "not all" translation here, as it more accurately reflects the most natural interpretation of the French sentence, i.e., with negation scoping over the universal. The "all not" translation is ambiguous in English.

[28]Unlike the negative universal sentences in Experiment 1, the universal quantifier here preceded the negation. Since we intended to use this condition as a scope ambiguity control for the homogeneity targets, we wanted to ensure that the sentences in the two conditions were as parallel as possible, in particular with respect to the relative surface order of the quantifier and negation.

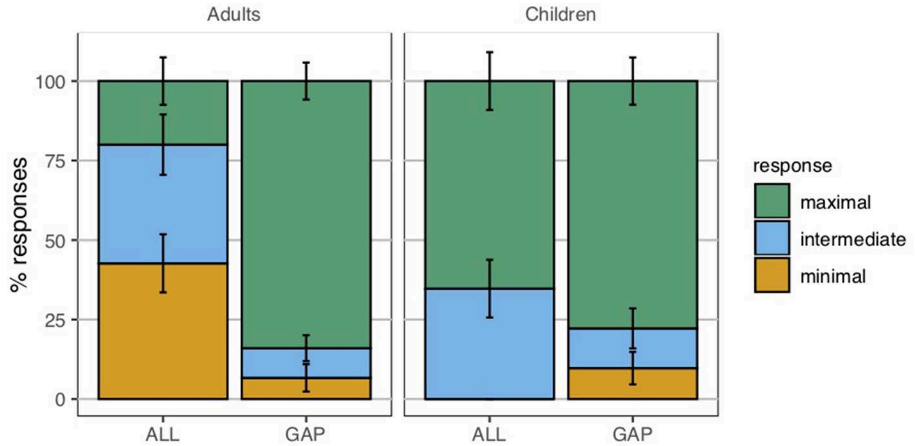


FIGURE 11 | Percentages of the reward types given in the existential *certains* "some" conditions. The ALL context corresponded to the scalar implicature targets, and the GAP context corresponded to the *incomplete description controls*. Minimal or intermediate rewards for existentially quantified sentences in ALL contexts were indicative of scalar implicatures. A less-than-maximal reward for existential descriptions of GAP contexts was indicative of incomplete description effects.

scalar implicature targets, i.e., existentially quantified sentences in ALL contexts, children gave more maximal rewards than adults, suggesting they computed fewer scalar implicatures than adults did. They also never gave *minimal* rewards on the basis of a false implicature and were thus, in a sense, more forgiving than adults. In response to the *incomplete description controls*, i.e., existentially quantified sentences in GAP contexts, children performed on a par with adults, generally maximally rewarding the puppet. This suggests that incomplete description effects do not play much of a role: children did not appear to be less inclined to give a high reward simply because the puppet had not described all of the objects in the picture.

5.2.2. Universal Quantification Conditions

Figure 12 displays the percentages of the reward types given in the universal quantification conditions. In response to the *partial truth controls*, i.e., positive universally quantified descriptions of GAP contexts, both adults and children gave less-than-intermediate rewards. The fact that children gave fewer minimal rewards than adults in this condition could be suggestive of a tendency to reward for partial truth.

In response to the *scope ambiguity controls*, i.e., negative universally quantified descriptions of GAP contexts, adults predominantly gave maximal rewards, which means that they interpreted the universal as scoping under negation. Children, on the other hand, were quite varied in their responses. While the maximal and minimal responses correspond to one of the readings of the sentence, intermediate responses may have two explanations. First, the intermediate rewards could reflect recognition of a sentence that may be construed as true, but is not unambiguously so. Second, some children could have accessed the surface scope interpretation of the negative sentences (*All of the hearts are such that they are not red*) and rewarded the partial truth of this sentence (which is literally false) with an intermediate response. Given the magnitude of the proportion of intermediate responses in this condition compared to the *partial truth* controls, however, it seems implausible that the latter could be solely responsible. Thus, it is plausible that scope ambiguity in itself would sometimes give rise to intermediate responses.

5.2.3. Plural Definite Description Conditions

Figure 13 displays the percentages of the reward types given in the plural definite description conditions. Adults and children generally performed as expected in three of the four unambiguous definite plural control conditions. In particular, they gave maximal rewards for the positive definite descriptions in ALL contexts (a *true* control) and minimal rewards for the positive definite descriptions in NONE contexts (a *false* control). They also gave minimal rewards for the negative definite descriptions in ALL contexts (a *false* control). In response to the negative definite descriptions in NONE contexts (a *true* control), however, children did not reward as maximally as adults did. A closer examination of children's responses and justifications suggests this was because children did not like the fact that the puppet's sentence mentioned a color that none of the objects shared. In other words, they may have seen some degree of infelicity associated with describing what color the objects were *not*, as opposed to what color they were[29].

As for the homogeneity targets, children and adults again differed in their treatment of plural definite descriptions in GAP contexts. First, as seen in **Figure 13**, adults generally gave the same responses to positive and negative homogeneity targets, while children tended to give greater rewards for positive homogeneity targets than for negative homogeneity targets. Second, the two groups differed in the distribution of individual

[29]It is worth noting that children were target-like on the same controls in Experiment 1. It's not entirely clear why the ternary task should bring out the infelicity of the negative descriptions more than the binary task. One possible reason is that choosing from three response options instead of two was more demanding, which could have pushed children to rely on a superficial strategy of quickly rejecting (or "punishing" the puppet for) any sentences that they perceived to be irrelevant. For example, they could have checked whether any of the objects in the picture matched the color mentioned by the puppet; if not, the puppet's statement would quickly be rejected as irrelevant or infelicitous.

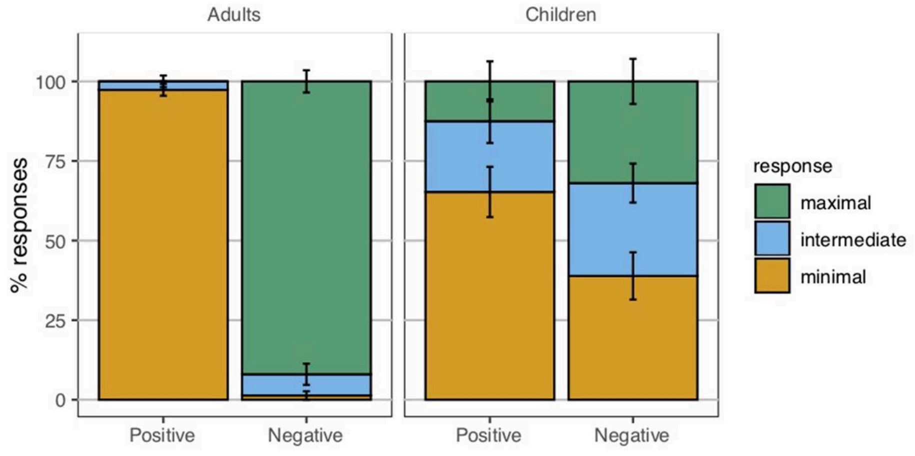


FIGURE 12 | Percentages of the reward types given in the universal *tous* "all" conditions. Positive universal descriptions of GAP contexts corresponded to *partial truth* controls; negative universal descriptions of GAP contexts corresponded to *scope ambiguity* controls. A greater-than-minimal reward for positive universal descriptions of GAP contexts was indicative of partial truth effects. A less-than-maximal reward for negative universal descriptions of GAP contexts was indicative of scope ambiguity effects.

participants across the different response categories. Participants were categorized as EXISTENTIAL if they gave the maximal reward on at least two of three positive target trials, and if they gave the minimal reward on at least two of three negative target trials. Participants were characterized as HOMOGENEOUS if they gave minimal or intermediate rewards on at least two of three positive and two of three negative target trials. Finally, participants were categorized as UNIVERSAL if they gave the minimal reward on at least two of three positive target trials, and if they gave the maximal reward on at least two of three negative target trials.

Table 6 represents the distribution of children and adults according to their performance on the homogeneity and scalar implicature targets[30]. Focusing first on the homogeneity targets, it is apparent that children and adults differed: while 23 of the 25 adults responded in a manner consistent with homogeneity, i.e., giving minimal or intermediate rewards to both positive and negative definite descriptions in GAP contexts, 12 children (mean age 5;08) displayed this adult pattern and 10 children (mean age 5;00) displayed the EXISTENTIAL response pattern, maximally rewarding the positive descriptions but minimally rewarding the negative descriptions ($\chi^2(2, N = 47) = 15.33, p < 0.001$). Two other children gave inconsistent responses.

Returning to the full distinctions presented in **Table 6**, we can discuss the individual responses to both homogeneity and scalar implicature targets together. We observe the same two subgroups of children as in Experiment 1: a subgroup of EXISTENTIAL children who failed to compute scalar implicatures, and a subgroup of HOMOGENEOUS children, only some of whom computed implicatures. As in Experiment 1, no child displayed the UNIVERSAL response pattern.

Finally, we took into account the *incomplete description*, *partial truth*, and *scope ambiguity* controls, in order to completely factor out these potential biases as described above. Recall that each participant received three repetitions of each kind of control. A participant was considered to have a bias against incomplete descriptions if they gave the maximal reward on fewer than two of the three trials. A participant was considered to have a bias in favor of partial truth if they gave the minimal reward on fewer than two of the three trials. Finally, a participant was considered to display a scope ambiguity effect if they gave the intermediate response on more than one of the three trials.

In **Table 7** we present the distribution of participants who passed this maximally conservative inclusion criterion. The remaining 21 adults and 9 children are those who we can be reasonably certain responded to the plural definite descriptions without any interfering or irrelevant biases. As was the case before the exclusions, we observe mostly homogeneous adults, and a homogeneous subgroup and an existential subgroup for children.

5.2.4. Non-randomness of Groupings

Since the ternary judgment task involves three response options, the number of logically possible groups, defined by how often a participant chose which option on which of the three relevant conditions (the two homogeneity targets THE-SOME-POS and THE-SOME-NEG, and the implicature target SOME-ALL-POS), is 27. Nevertheless, 22 of the 24 children fall into only three of these groups, and it is precisely the groups which, from a theoretical point of view, correspond to the three groups in which 23 of the 24 children were found in Experiment 1. It is thus highly unlikely that the five participants in the homogeneous/−implicature group are there simply by virtue of giving random responses. Since the relevant p-values are guaranteed to be much lower

[30]Again, some would prefer an analysis that does not bin participants into categories, but we include this discussion here to explore the possible interpretive preferences observable in our two participant groups; as in Experiment 1, participants turned out to align remarkably well into a subset of the possible categories.

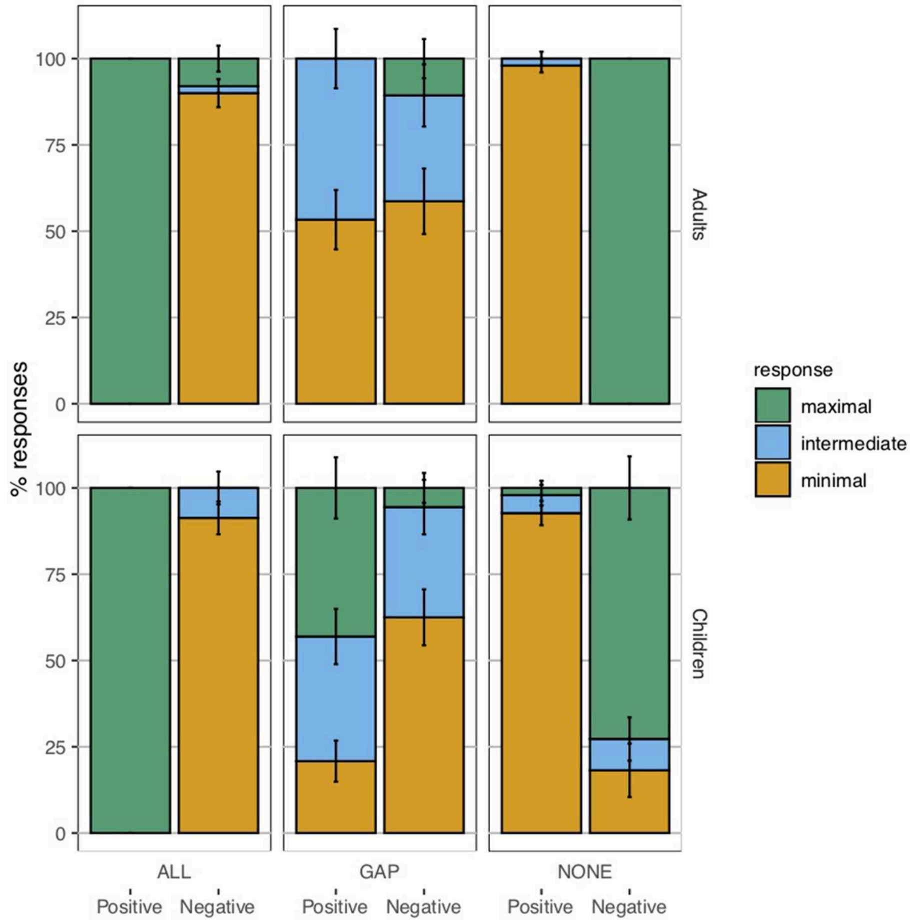


FIGURE 13 | Percentages of the reward types given in the definite *les* "the" conditions. True controls corresponded to positive plural definite descriptions of ALL contexts and negative plural definite descriptions of NONE contexts. False controls corresponded to positive plural definite descriptions of NONE contexts and negative plural definite descriptions of ALL contexts. Homogeneity targets corresponded to plural definite descriptions of GAP contexts.

TABLE 6 | Distribution of participants across response types, according to performance on homogeneity and scalar implicature targets.

	Adults		Children	
	− Implicature	+ Implicature	− Implicature	+ Implicature
Homogeneous	2	21	5	7
Existential	0	0	10	0
Universal	2	0	0	0

Two children gave inconsistent responses and are not included in the table.

TABLE 7 | Distribution of participants across response types, after applying a maximally stringent exclusion criterion that eliminated any participants who could potentially have had a bias for partial truth, against incomplete description, or against scope ambiguity.

	Adults		Children	
	− Implicature	+ Implicature	− Implicature	+ Implicature
Homogeneous	2	18	2	2
Existential	0	0	5	0
Universal	1	0	0	0

than even for Experiment 1 (section 4.2.3), we do not calculate them here.

5.2.5. Improved Group Assignment

The purpose of this section is, again, to obtain a quantitative assessment of the preceding characterization of the data in terms of assigning children to groups. The question we are interested in is whether there is evidence for the existence of children with truly homogeneous interpretations for definite plurals but who do not compute implicatures. To this end, we will describe an analysis that allows us to decide for each participant whether they have implicatures and what reading they assign to the definite plural. The possible readings for the definite plurals that we consider are the following:

EXI Low-scope existential interpretation for definite plurals (as before).

HOM Truly homogeneous interpretation, which should lead to an intermediate response in both positive and negative THE-GAP conditions.

TABLE 8 | Predicted responses to each condition ([determiner]-[context]-[polarity]) for each of the possible groups.

	EXI		HOM		SA		WS		PT	
	+SI	-SI	+SI	-SI	+SI	-SI	+SI	-SI	+SI	-SI
ALL-GAP-NEG	1	1	1	1	1	1	0	0	0	0
SOME-GAP-POS	1	1	1	1	0	0	1	1	1	1
THE-ALL-NEG	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
THE-GAP-NEG	−1	−1	0	0	−1	−1	0	0	0	0
THE-NONE-POS	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
THE-GAP-POS	1	1	0	0	1	1	1	1	0	0
THE-NONE-NEG	1	1	1	1	1	1	1	1	1	1
THE-ALL-POS	1	1	1	1	1	1	1	1	1	1
SOME-ALL-POS	0	1	0	1	0	1	0	1	0	1
ALL-GAP-POS	−1	−1	−1	−1	−1	−1	−1	−1	0	0

SA Universal interpretation with scope ambiguity effects, which should lead to a minimal response in the THE-GAP-POS condition since there is no possibility for scope ambiguity here, but to an intermediate response in the THE-GAP-NEG condition because the sentence is either true or false depending on where the universal takes scope with respect to negation.

WS A strictly wide-scope universal interpretation, which should yield minimal responses in both positive and negative THE-GAP conditions.

PT Wide-scope universal interpretation with partial truth effects, which should yield intermediate responses in both THE-GAP conditions, like HOM, but should additionally yield an intermediate response in the ALL-GAP-POS condition (where HOM would yield a minimal response).

We thus obtain in principle 5(EXI,HOM,SA,WS,PT) × 2(+SI,−SI) possible groups of participants. Each of these groups corresponds to a unique pattern of responses to the different conditions, as described in **Table 8**. The upcoming analyses fit ordinal regression models which assign each participant to a given group, given this participant's actual responses[31]. The models may allow for different groups to be considered, and in order to decide whether it is meaningful to say that some participants belong to a particular group, we ask whether models that include that group are superior to models that do not include that group, all else being equal. One problem is that it is not necessarily possible to reliably estimate the relevant models with the whole dataset while considering all possible groups at once, so below we propose several analyses which are essentially similar but differ in what assumptions they rely on to simplify this computational limitation. In all these analyses, we rely on the results of Experiment 1, where only a single child was categorized as belonging to the group EXI/+SI, in not including that group in any of the models.

Analysis 1: no partial truth, implicatures imply homogeneity

In this analysis, we restricted the dataset to the conditions with THE plus the implicature-relevant condition SOME-ALL-POS. There is little reason to think that partial truth was playing any role and, accordingly, this analysis does not consider the possibility of a PT group (the role of PT groups is evaluated separately in Analysis 3). Furthermore, we assume here that every child who had acquired implicatures had also reached an adult-like stage for homogeneity. The only +SI group allowed in these models was thus the HOM/+SI group. Apart from these restrictions, the models in this analysis explore all combinations of HOM/−SI, SA/−SI, and WS/−SI[32].

Table 9 shows the estimated log pointwise predictive likelihoods (elpd) and their standard errors for each of these models. Overall, models that included the HOM/−SI group were superior to those that did not, providing evidence in favor of

[31] The models were thus similar to the ones used in the previous experiment, except that we performed ordinal regression because the responses could now take 3, and not only 2, values. The analysis was again performed with JAGS/`rjags` (Plummer, 2003). The prior for the two threshold parameters in the ordinal regression was set to a normal distribution with mean 0 and precision 0.001, while the prior for the slope parameter was set to the positive half of the same distribution. The prior over the group parameter γ was uniform (over those groups which were available to the model). Leave-one-out cross-validation was performed as before on the basis of 5,000 samples after 5,000 burn-in iterations from 4 chains, for a total of 20,000 samples per model and data point.

[32] Since HOM and PT do not differ on the conditions under consideration (but only on ALL-GAP-POS, which was not included), the latter group was excluded from the models. Note that while we did take into account the possibility of scope ambiguity effects, we did not use the ALL-GAP-NEG data points to estimate their prevalence. The effect of the scope ambiguity of the negative universal sentences was very strong in that all possible responses were chosen in a sizeable percentage of cases. Given that this is not so for negative definite description sentences, these effects are clearly more prevalent with the universal *all*, and we found that the scope ambiguity effect with *all* dominated the group assignment choices of the models at the expense of an accurate categorization on the basis of the definite description data. In other words, the model would rather give up accuracy on the definite description data than miss the scope ambiguity effects with *all*. Furthermore, there was no a priori reason to expect the scopal behavior of different quantifiers (in this case, definite descriptions and *all*) to be the same. We thus concluded that the inclusion of these data points would hamper, rather than improve, the analysis.

TABLE 9 | Results of leave-one-out cross-validation for Experiment 2.

			Analysis 1		Analysis 2		An. 1 *vs*. An. 2		Analysis 3	
HOM/−SI	WS/−SI	SA/−SI	elpd	(se)	elpd	(se)	Δ_{elpd}	(se)	Δ_{elpd}	(se)
✓	✓		−238.9	(18.0)	−239.6	(18.2)	0.8	(3.6)	8.9	(4.3)
✓		✓	−240.6	(17.6)	−244.8	(17.8)	4.1	(3.8)	8.8	(4.5)
✓	✓	✓	−241.4	(17.9)	−241.6	(18.2)	0.3	(3.6)	8.3	(4.4)
✓			−241.5	(17.5)	−249.0	(17.7)	7.5	(4.6)	9.9	(4.7)
	✓		−247.7	(18.6)	−249.3	(18.7)	1.6	(3.8)	1.2	(6.9)
	✓	✓	−248.7	(18.5)	−249.8	(18.7)	1.1	(3.7)	1.6	(6.8)
		✓	−256.5	(18.2)	−253.0	(18.2)	−3.5	(5.2)	1.6	(6.9)

All models included groups HOM/+SI and EXI/-SI on top of those marked in the table. Models in Analysis 2 additionally included SA/+SI and WS/+SI. Models in Analyses 1 and 2 were fitted on THE conditions plus SOME-ALL-POS. Models in Analysis 3 were fitted on the same data plus ALL-GAP-POS. The column for Analysis 3 shows the comparison of models that included the PT/−SI with models that did not (positive numbers favor the latter).

the existence of a group of children with access to homogeneous readings but not to implicatures[33].

Analysis 2: no partial truth, no assumption that implicatures imply homogeneity

This analysis differed from Analysis 1 in that it did not assume that implicatures imply homogeneous readings; that is, the groups SA/+SI and WS/+SI were systematically included in all models as possible groups a child could fall in. The overall picture remains largely the same, with elpds in the same range as in Analysis 1 (**Table 9**), and favoring models making use of the HOM/−SI group.

A comparison of the models from Analysis 2 to the corresponding models from Analysis 1 (also provided in **Table 9**) reveals that those from Analysis 1 actually perform better, suggesting that the assumption in Analysis 1 that homogeneity is systematically acquired earlier than implicatures is warranted[34,35].

Analysis 3: the role of partial truth

For Analysis 3, we are interested in evaluating the role of the partial truth strategy. The target is thus the comparison of models with and without PT groups. Given the results of Analysis 2, we start over from Analysis 1, assuming that homogeneity precedes implicatures, i.e., dropping all +SI group except HOM/+SI. In Analysis 3, the condition ALL-GAP-POS was included alongside the conditions used in Analysis 1, because it is now necessary to differentiate the newly added PT/−SI group from the HOM/−SI group. The last column of **Table 9** presents a comparison of models with a partial truth PT/−SI group with the corresponding models without such a group. The comparison uniformly comes out in favor of the models without PT/−SI. Hence, this analysis provides no evidence for the existence of the PT/−SI group or, to put it differently, in favor of the partial truth strategy.

5.3. Summary

The results of Experiment 2 replicate the essential findings of Experiment 1 insofar as, if one were to collapse intermediate and minimal rewards in the ternary paradigm, the resulting picture is very similar to what we saw in Experiment 1 on all the crucial points. Furthermore, we find that even if some of the children who do not compute implicatures may have a wide-scope universal reading for the definite plural (which Experiment 1 could not distinguish from a truly homogeneous one), there is evidence for a group with homogeneous readings and, nonetheless, no implicatures.

6. DISCUSSION

The results of our experiments revealed, by and large, three groups of children. The first group of children did not compute implicatures and interpreted definite plurals as existentials (that scope under negation). A question that is raised by this state of affairs is the following. Children are evidently able to reach truth conditions equivalent to those of adults for negated sentences by recognizing that definite plurals, interpreted existentially, have to scope under negation. But why would they hypothesize an existential meaning in the first place when it results in truth conditions for affirmative sentences that are so different from those of adults? We can only offer some speculation as to how this asymmetry might come about. It is well-known that sentences with definite plurals are not infrequently used when there are some exceptions, even though under scrutiny we would not judge such sentences as strictly true. This phenomenon is known as *non-maximality* (Brisson, 1998; Lasersohn, 1999; Malamud,

[33]The top model also included the WS/−SI group. It should, however, be noted that the models are likely to overestimate the prevalence of the WS/−SI group. Both our own adult data and the data from Križ and Chemla (2015) suggest that adults often judge sentences as false when they are really undefined due to a homogeneity violation. Adults do this more often than they judge the same sentences true, and also more often than they judge a sentence with a false implicature as false (cf. **Figures 12**, **13** above). Our simple ordinal regression models do not account for this fact and therefore categorize children into the WS/−SI group even when they are really in the HOM/−SI group, simply translating the underlying undefined status of the sentence to a minimal response.

[34]The only exception was SA/−SI, which was independently the worst model. The reason why SA/−SI is a bit better in Analysis 2 is presumably that it now has a way of assigning children to a WS/−SI group by sacrificing fit on the implicature data: (some) children that ought to be in WS/−SI were instead categorized as WS/+SI. This is not possible in Analysis 1, where no WS/+SI group exists.

[35]If we look at how the maximal model categorizes children (by maximal posterior probability of group), we find that only one child is assigned to SA/+SI and no child is assigned to WS/+SI, whereas there are six participants in the HOM/+SI group. In light of this model's failure to perform markedly better than more parsimonious ones, we conclude that it is probably overfitted and that the assumption that implicatures imply homogeneity need not be given up on the basis of these results.

2012; Križ, 2015b). An example from Lasersohn (1999) is (26), which can be felicitously used to describe a situation in which there are nevertheless a few insomniacs who are reading in bed and not actually asleep.

(26) The townspeople are asleep.

While the exceptions that can be ignored by way of non-maximality are typically few in number, in the right contexts, non-maximal readings can effectively turn existential, such as in this example from Malamud (2012):

(27) Context: *Mary has a large house with over a dozen windows in different rooms. She locks up and leaves to go on a road trip with her friend Max, forgetting to close just a few of the many windows in various rooms. A few minutes into the ride, Max says, "There is a thunderstorm coming. Is the house going to be OK?" Mary replies:*

Oh my, we have to go back — the windows are open!

Assuming that young children do not have the interpretive mechanisms available to simultaneously make sense of homogeneity and non-maximality, it might be reasonable for them to assign an existential interpretation to the definite plural in order to be able to accommodate such non-maximal uses. There is reportedly an asymmetry in the availability of non-maximal readings for affirmative and negated sentences, possibly related to the kinds of contexts in which we would use them (Križ, 2015b). If this is correct, then children will observe much fewer non-maximal readings of negated sentences, which could lead them to assume that such sentences are indeed only false when the predicate holds of none of the individuals in question. This, they can accommodate by assuming that the existentially interpreted definite plural has to take scope under negation[36].

A second group of children was found to have already acquired scalar implicatures as well as a homogeneous interpretation of plural definite descriptions, and was therefore adult-like.

Finally, a third group of children appeared to access the homogeneous interpretation of the plural definite descriptions without computing scalar implicatures. A closer look in Experiment 2 suggests that some of these children actually assign a wide-scope universal interpretation to the definite plural. This would seem to be a natural hypothesis on the part of these children[37], since, setting non-maximality aside, the data that are needed to distinguish this hypothesis from the correct homogeneous reading (e.g., involving definite plurals in the scope of non-monotonic quantifiers) are quite subtle and presumably not all too frequent in the speech children are exposed to. Importantly, however, there is still evidence for a group of children who do assign adult-like homogeneous readings to definite plurals while not computing scalar implicatures.

Given that (at least some) children start out with an existential meaning for definite plurals, and that by the time they have acquired scalar implicatures, they have also reached an adult-like homogeneous meaning for definite plurals, it is tempting to think that implicatures are, in fact, the way *by which* they obtain such a homogeneous meaning. This would accord exactly with Magri's (2014) implicature-based theory of homogeneity, in which definite plurals are assumed to have an existential literal meaning.

While it cannot be excluded that some children transition to the adult-like state via the implicature theory of homogeneity, our data provide evidence that the implicature theory is not a correct description of the adult state itself. Since the implicature theory requires the implicature from *some* to *not all* as a subcomputation of the implicature that is behind homogeneity effects, it predicts that homogeneous readings should not be more frequent than this scalar implicature. This is inconsistent with our adult data. If, however, as our data indicate, the implication between scalar implicatures and homogeneity is only unidirectional even in children (so that there are children with homogeneous definite plurals but no implicatures), it is also not clear that the implicature theory has a role to play in development. Rather, it seems quite plausible that the two phenomena are independent and that homogeneity (whatever its proper analysis) is simply acquired earlier than scalar implicatures[38].

7. CONCLUSION

In this paper, we presented two experiments that tested children's interpretation of sentences containing plural definite descriptions, such as the affirmative *The trucks are blue* and the negated *The trucks are not blue*. These experiments also included testing children's ability to compute scalar implicatures, and therefore allowed us to directly compare children's performance on the two phenomena. This in turn afforded us the opportunity to assess the viability of scalar implicature accounts of homogeneity.

The data from our experiments confirm previous findings (Karmiloff-Smith, 1979; Caponigro et al., 2012) that (many) children interpret definite plurals as existential, and extend this existential interpretation to the context of negation, where we find that the existential takes low scope. This corresponds to the literal meaning hypothesized by the implicature theory of homogeneity (Magri, 2014). However, the finding of children (and adults) who have access to homogeneity while failing to compute the scalar implicature that is argued to be a sub-computation of homogeneity is incompatible with the predictions of this theory. While we have remained agnostic as to the nature of homogeneity in the adult grammar, our experiments suggest that it is a phenomenon distinct from scalar implicatures and acquired earlier by children.

DATA AVAILABILITY STATEMENT

The data and R scripts for this study are available online at: http://semanticsarchive.net/Archive/DM5YjA1M/Tieu-Kriz-Chemla-AcqHomogeneity.html.

[36]Note that a low-scope universal reading for definite plurals is, in light of this input, an implausible hypothesis, such that its absence in children is not surprising.

[37]Indeed, Caponigro et al.'s (2012) sample of plural definite descriptions in child-directed speech is entirely associated with maximal interpretations.

[38]The ages of the groups in Experiment 1 show a trend toward children without scalar implicatures being younger than those with scalar implicatures, but the difference is not statistically significant in our sample, which was drawn from the rather narrow range of 4; 04, 14 − 5; 03, 24. The main argument is, of course, the unidirectional implication from scalar implicatures to homogeneity that we seem to observe.

ETHICS STATEMENT

Ethical approval for this study was obtained from the CERES (Comité d'évaluation éthique des projets de recherche en santé non soumis à CPP) under approval number 2013/46.

AUTHOR CONTRIBUTIONS

LT, MK, and EC conceived and designed the study. LT prepared and carried out the experiments. MK performed the statistical analysis. All authors contributed to writing, revising, reading, and approving the submitted manuscript.

FUNDING

The research leading to this work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n.313610, and by ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.

ACKNOWLEDGMENTS

For helpful feedback and discussion, we would like to thank Alexandre Cremers, Ewan Dunbar, Napoleon Katsos, Tal Linzen, Jacopo Romoli, and Jérémy Zehr. We would also like to thank Monica Barbir for her assistance with stimuli and piloting, Anne-Caroline Fiévet for coordinating testing at local preschools, Charlotte Hauser for her role in recording the audiovisual stimuli, and Florian Pellet for implementing the adult version of the task. We are grateful to the families of the children who participated in our study, as well as the preschool directors and teachers in Paris, who graciously allowed us to conduct our study in their schools.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02329/full#supplementary-material

REFERENCES

Bar-Lev, M. (2018). *Free choice, homogeneity, and innocent inclusion* (PhD thesis), Hebrew University of Jerusalem, Jerusalem, Israel.

Barner, D., Brooks, N., and Bale, A. (2011). Accessing the unsaid: the role of scalar alternatives in children's pragmatic inferences. *Cognition* 188, 87–96. doi: 10.1016/j.cognition.2010.10.010

Barner, D., Thalwitz, D., Wood, J., and Carey, S. (2007). On the relation between the acquisition of singular-plural morpho-syntax and the conceptual distinction between one and more than one. *Dev. Sci.* 10, 365–373. doi: 10.1111/j.1467-7687.2007.00591.x

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Berko, J. (1958). The child's learning of English morphology. *Word* 14, 150–177.

Braine, M. D. S., and Rumain, B. (1981). Children's comprehension of 'or': evidence for a sequence of competencies. *J. Exp. Child Psychol.* 31, 46–70.

Breheny, R. (2005). "Exhaustivity, homogeneity, and definiteness," in *Proceedings of the Fifth Amsterdam Colloquium*, eds P. Dekker, and M. Franke (Amsterdam), 59–65.

Brisson, C. (1998). *Distributivity, maximality, and floating quantifiers* (PhD thesis), New Brunswick, NJ: Rutgers University.

Büring, D., and Križ, M. (2013). It's that and that's it! exhaustivity and homogeneity presuppositions in clefts (and definites). *Semant. Pragmat.* 6, 1–29. doi: 10.3765/sp.6.6

Caponigro, I., Pearl, L., Brooks, N., and Barner, D. (2010). "On the acquisition of maximality in free relative clauses and plural definite descriptions," in *Proceedings of Semantics and Linguistic Theory (SALT) 20,* eds N. Li and D. Lutz (Ithaca, NY: CLC Publications, Cornell University), Vol. 20, 508–524. doi: 10.3765/salt.v20i0.2564

Caponigro, I., Pearl, L., Brooks, N., and Barner, D. (2012). Acquiring the meaning of free relative clauses and plural definite descriptions. *J. Semant.* 29, 261–293. doi: 10.1093/jos/ffr014

Chierchia, G., Crain, S., Guasti, M. T., and Thornton, R. (2001). "'Some' and 'or': a study on the emergence of logical form," in *Proceedings of the 24th Boston University Conference on Language Development*, eds S. C. Howell, S. A. Fish, and T. Keith-Lucas (Somerville, MA: Cascadilla Press), 22–44.

Chierchia, G., Fox, D., and Spector, B. (2011). "The grammatical view of scalar implicatures and the relationship between semantics and pragmatics," in *An International Handbook of Natural Language Meaning Volume 3*, eds C. Maienborn, K. von Heusinger, and P. Portner (Berlin: Mouton de Gruyter).

Conroy, S., Lidz, J., and Musolino, J. (2009). The fleeting isomorphism effect. *Lang. Acquisit.* 16, 106–117. doi: 10.1080/10489220902769242

Crain, S., and Thornton, R. (1998). *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, MA: MIT Press.

Davies, B., Xu Rattanasone, N., and Demuth, K. (2016). Two-year-olds' sensitivity to inflectional plural morphology: allomorphic effects. *Lang. Learn. Dev.* 13, 38–53. doi: 10.1080/15475441.2016.1219257

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., and Pethick, S. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 1–173. doi: 10.2307/1166093

Fodor, J. D. (1970). *The linguistic description of opaque contexts*. (PhD thesis), Massachusetts Institute of Technology, Cambridge, MA, United States.

Fox, D. (2007). "Free choice and the theory of scalar implicatures," in *Presupposition and Implicature in Compositional Semantics*, eds U. Sauerland, and P. Stateva (Basingstoke: Palgrave Macmillan), 71–120.

Gajewski, J. (2005). *Neg-raising: polarity and presupposition* (PhD thesis), Cambridge, MA: MIT.

Grice, P. (1975). "Logic and conversation," in *The Logic of Grammar*, eds D. Davidson, and G. H. Harman (Encino, CA: Dickenson Publishing Company), 64–75.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., and Guasti, M. T. (2001). "At the semantics/pragmatics interface in child language," in *Proceedings of Semantics and Linguistic Theory 11*, eds R. Hastings, B. Jackson, and Z. Zvolenszky (Ithaca, NY: Cornell University, CLC Publications), 231–247.

Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge Studies in Linguistics 24. Cambridge: Cambridge University Press.

Katsos, N., and Bishop, D. V. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition* 120, 67–81. doi: 10.1016/j.cognition.2011.02.015

Križ, M. (2015a). *Aspects of homogeneity in the semantics of natural language* (PhD thesis), University of Vienna, Vienna, Austria.

Križ, M. (2015b). Homogeneity, non-maximality, and 'all'. *J. Semant.* 33, 493–539. doi: 10.1093/jos/ffv006

Križ, M., and Chemla, E. (2015). Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Nat. Lang. Semant.* 23, 205–248. doi: 10.1007/s11050-015-9114-z

Križ, M., and Spector, B. (2017). *Interpreting Plural Predication: Homogeneity and Non-Maximality*. Manuscript, Institut Jean-Nicod.

Lasersohn, P. (1999). Pragmatic halos. *Language* 75, 522–551.

Lewis, S., Hacquard, V., and Lidz, J. (2017). "Think" pragmatically: children's interpretation of belief reports. *Lang. Learn. Dev.* 13, 395–417. doi: 10.1080/15475441.2017.1296768

Link, G. (1983). "The logical analysis of plurals and mass terms: a lattice-theoretical approach," in *Meaning, Use and Interpretation of Language*, eds R. Baeuerle, C. Schwarze, and A. V. Stechow (Berlin: DeGruyter), 302–323.

Löbner, S. (1987). "The conceptual nature of natural language quantification," in *Proceedings of the '87 Debrecen Symposium on Logic and Language*, eds I. Rusza, and A. Szabolcsi (Budapest: Akadémiai Kiadó).

Löbner, S. (2000). Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguist. Philos.* 23, 213–308. doi: 10.1023/A:1005571202592

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edn.* (Mahwah, NJ: Lawrence Erlbaum Associates).

Magri, G. (2014). "An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening," in *Pragmatics, Semantics and the Case of Scalar Implicatures*, ed S. Pistoia Reda (London: Palgrave Macmillan), 99–145.

Malamud, S. (2012). The meaning of plural definites: a decision-theoretic approach. *Semant. Pragmat.* 5, 1–58. doi: 10.3765/sp.5.3

Maratsos, M. (1974). Preschool children's use of definite and indefinite articles. *Child Dev.* 45, 446–455.

Maratsos, M. (1976). *The Use of Definite and Indefinite Reference in Young Children*. Cambridge: Cambridge University Press.

Marchman, V., Plunkett, K., and Goodman, J. (1997). Overregularization in English plural and past tense inflectional morphology: a response to Marcus (1995). *J. Child Lang.* 24, 767–779.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T., and Xu, F. (1992). *Overregularization in Language Acquisition, Volume 57 of Monographs of the Society for Research in Child Development*. Hoboken, NJ: Wiley.

Matthewson, L., Bryant, T., and Roeper, T. (2001). "A Salish stage in the acquisition of English determiners: unfamiliar 'definites'," in *Proceedings of SULA: The Semantics of Under-Represented Languages in the Americas, Volume 25 of University of Massachusetts Occasional Papers in Linguistics* (Amherst, MA: Graduate Linguistics Student Association (GLSA), Department of Linguistics, University of Massachusetts).

Mervis, C. B., and Johnson, K. E. (1991). Acquisition of the plural morpheme: a case study. *Dev. Psychol.* 27, 222–235.

Modyanova, N., and Wexler, K. (2008). "Maximal trouble in free relatives," in *Proceedings of the 32nd Annual Boston University Conference on Language Development*, eds H. Chan, H. Jacob, and E. Kapia (Boston, MA), 287–298.

Munn, A., Miller, K., and Schmitt, C. (2006). "Maximality and plurality in children's interpretation of definites," in *Proceedings of the 30th Annual Boston University Conference on Child Language Development*, eds D. Bamman, T. Magnitskaia, and C. Zaller (Somerville, MA: Cascadilla Press), 377–387.

Musolino, J., and Lidz, J. (2006). Why children aren't universally succesful with quantification. *Linguistics* 44, 817–852. doi: 10.1515/LING.2006.026

Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition* 78, 165–188. doi: 10.1016/S0010-0277(00)00114-1

Noveck, I. A., and Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain Lang.* 85, 203–210. doi: 10.1016/S0093-934X(03)00053-1

Papafragou, A., and Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition* 86, 253–282. doi: 10.1016/S0010-0277(02)00179-8

Pérez-Leroux, A. T., Munn, A., Schmitt, C., and DeIrish, M. (2004). Learning definite determiners: genericity and definiteness in English and Spanish. in *Supplement to the Proceedings of the 28th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Plummer, M. (2003). "JAGS: a program for analysis of Bayesian graphical models using gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, eds K. Hornik, F. Leisch, and A. Zeileis (Vienna: Technische Universität Wien).

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Royle, P., Fromont, L. A., and Drury, J. E. (2018). Definiteness and maximality in French language acquisition, more adult-like than you would expect. *Front. Commun.* 3:27. doi: 10.3389/fcomm.2018.00027

Sauerland, U., Andersen, J., and Yatsushiro, K. (2005). "The plural is semantically unmarked," in *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, eds S. Kepser, and M. Reis (Berlin: Mouton de Gruyter), 413–434.

Schaeffer, J., and Matthewson, L. (2005). Grammar and pragmatics in the acquisition of article systems. *Nat. Lang. Linguist. Theory* 23, 53–101. doi: 10.1007/s11049-004-5540-1

Schafer, R., and de Villiers, J. (2000). "Imagining articles: what a and the can tell us about the emergence of DP," in *Proceedings of the 24th Annual Boston University Conference on Language Development*, eds S. C. Howell, S. A. Fish, and T. Keith-Lucas (Somerville, MA: Cascadilla Press), 609–620.

Schmerse, D., Lieven, E., and Tomasello, M. (2014). Young children use shared experience to interpret definite reference. *J. Child Lang.* 42, 1146–1157. doi: 10.1017/S0305000914000555

Schwarz, F. (2013). "Maximality and definite plurals – experimental evidence," In *Proceedings of Sinn und Bedeutung 17*, eds E. Chemla, V. Homer, and G. Winterstein (Paris: Ecole Normale Supérieure), 509–526.

Schwarzschild, R. (1994). Plurals, presuppositions and the sources of distributivity. *Nat. Lang. Semant.* 2, 201–248.

Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., and Fox, D. (2016). Children interpret disjunction as conjunction: consequences for theories of implicature and child development. *Nat. Lang. Semant.* 24, 305–352. doi: 10.1007/s11050-016-9126-3

Skordos, D., and Papafragou, A. (2016). Children's derivation of scalar implicatures: alternatives and relevance. *Cognition* 153, 6–18. doi: 10.1016/j.cognition.2016.04.006

Spector, B. (2013). "Homogeneity and plurals: from the strongest meaning hypothesis to supervaluations," in *Sinn und Bedeutung 18*. Vitoria-Gasteiz: University of the Basque Country (UPV/EHU).

Stiller, A., Goodman, N., and Frank, M. (2015). *Ad-hoc* implicature in preschool children. *Lang. Learn. Dev.* 11, 176–190. doi: 10.1080/15475441.2014.927328

Tieu, L., Bill, C., Romoli, J., and Crain, S. (2014). "Plurality inferences are scalar implicatures: evidence from acquisition," in *Proceedings of the 24th Semantics and Linguistic Theory Conference*, eds T. Snider, S. D'Antonio, and M. Weigand (LSA and CLC Publications), 122–136.

Tieu, L., and Lidz, J. (2016). Beyond NPI licensing: children's knowledge of the semantics of 'any'. *Lang. Acquisit.* 23, 311–332. doi: 10.1080/10489223.2016.1176172

Tieu, L., Romoli, J., Zhou, P., and Crain, S. (2016). Children's knowledge of free choice inferences and scalar implicatures. *J. Semant.* 33, 269–298. doi: 10.1093/jos/ffv001

Tieu, L., Yatsushiro, K., Cremers, A., Romoli, J., Sauerland, U., and Chemla, E. (2017). On the role of alternatives in the acquisition of simple and complex disjunctions in French and Japanese. *J. Semant.* 34, 127–152. doi: 10.1093/jos/ffw010

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27, 1413–1432.

Winitz, H., Sanders, R., and Kort, J. (1981). Comprehension and production of the /schwa z/ plural allomorph. *J. Psycholinguist. Res.* 10, 259–271.

Wood, J. N., Kouider, S., and Carey, S. (2009). Acquisition of singular-plural morphology. *Dev. Psychol.* 45, 202–206. doi: 10.1037/a0014432

Zapf, J. A., and Smith, L. B. (2008). Meaning matters in children's plural productions. *Cognition* 108, 466–476. doi: 10.1016/j.cognition.2008.03.008

Zhou, P., Romoli, J., and Crain, S. (2013). "Chidren's knowledge of alternatives," in *Proceedings of Semantics and Linguistic Theory (SALT) 23*, ed T. Snider (Ithaca, NY: Cornell University, CLC Publications), 632–651.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 July 2020
doi: 10.3389/fpsyg.2020.01274

Generics and Alternatives

Arnold Kochari†, Robert Van Rooij and Katrin Schulz**

Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands

In this paper we argue that for the (probabilistic) interpretation of generic sentences of the form "*G*s are *f*," three types of alternatives play a role: (i) alternative features of *f*, (ii) alternative groups, or kinds, of *G*, and (iii) alternative causal background factors. In the first part of this paper we argue for the relevance of these alternatives. In the second part, we describe the results of some experiments that empirically tested in particular the second use of alternatives.

Keywords: generics, alternatives, probabiltiy, semantics, experiments

Edited by:
Katharina Spalek,
Humboldt University of Berlin, Germany

Reviewed by:
Bernhard Nickel,
Harvard University, United States
Olav Mueller-Reichau,
Leipzig University, Germany

****Correspondence:***
Robert Van Rooij
r.a.m.vanrooij@uva.nl
Katrin Schulz
k.schulz@uva.nl

†Authors are listed in alphabetical order

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *14 August 2019*
Accepted: *15 May 2020*
Published: *03 July 2020*

Citation:
Kochari A, Van Rooij R and Schulz K (2020) Generics and Alternatives. Front. Psychol. 11:1274. doi: 10.3389/fpsyg.2020.01274

1. INTRODUCTION

Bare plural (or BP) generic sentences like "Birds fly" and "Tigers are striped" (which we take to have the form "*G*s are *f*") are sentences that, by their very nature, express useful generalizations. Accounting for the meaning of these sentences has been proven to be notoriously difficult. The problem is to account for the fact that generics allow for exceptions. We believe that birds fly, even though not all birds do or can fly.

One very popular solution to this problem proposed in the linguistic literature is to assume the presence of a generic operator, which is then analyzed as a universal quantifier with a restricted domain of quantification: for the generic to be true all the *relevant* or *normal* members of the group *G*, or all the members under *normal* circumstances, have to have the feature *f* under discussion (e.g., all relevant or normal birds fly, or all objects being birds under normal circumstances fly) (cf. Asher and Morreau, 1995). But without an independent and satisfying account of what relevance and normalcy is this will not bring us any closer to a true solution of the problem.

We will follow here a different line of approach to the meaning of generic sentences. This is the idea that their meaning should be related to the frequency with which we observe a member of the group *G* to bear feature *f*. A very natural and often explored approach along these lines is the majority rule for the interpretation of generics (Cohen, 1999, 2004). According to the majority rule a generic is true in case the probability of a member of group *G* having feature *f* is high, (much) higher than $\frac{1}{2}$.

Definition 1. *A simple majority rule for generics.*
A generic sentence 'Gs are f' is true in case $P(f|G) > \frac{1}{2}$.

Thus, taking a generic like (1), according to definition 1 this sentence is true in case the majority of the birds fly.

(1) Birds fly.

This natural approach to the meaning of BP generics nicely accounts for the fact that not all birds need to fly in order for the generic to be true and still plays an important role in the literature on generic expressions. But while it has been shown that frequency does play a role for the meaning of generics (e.g., Prasada and Dillingham, 2006), this approach has difficulties to account for the different degree with which generics allow for exceptions. In some cases we are willing to accept generic sentences even in cases where only very few group members carry the feature in question. For instance, a generic statement like (2) is generally accepted to be true, even though only 1% of mosquitoes are actually carriers of the virus (Cox, 2004).

(2) Mosquitoes carry the West Nile virus.

There are many more studies that enforce the conclusion that the truth of a generic sentence cannot be in general reduced to a high conditional probability of f on G. Experimental evidence was first provided by Gilson Gilson and Abelson (1965), but similar conclusions also emerged in the linguistic literature (e.g., Lawler, 1973; Dahl, 1975; Carlson, 1977; Declerck, 1986). These results were then confirmed in psychological studies (e.g., Prasada, 2000; Gelman, 2004; Gelman and Bloom, 2007; Cimpian et al., 2010a).

Especially in the psychological literature on generics the observation that the meaning of generics cannot be reduced to a high conditional probability has then been taken to show that there is no systematic relation between the meaning of generics and statistical information (e.g., Leslie, 2008; Cimpian et al., 2010b). This conclusion is wrong in our eyes, or at least premature. The fact that Rule 1 is not an adequate description of the truth conditions of generic sentences does not show that no statistically based rule for the meaning of generics is possible—as claimed by the authors mentioned above. More concretely, in this paper we will show that if we take into account alternatives we can substantially improve on Rule 1 and can account for examples like (2).

We will argue for an extension of this rule involving three sets of alternatives.

1. Alternatives of the property f, $Alt(f)$, limit the domain of the probability function involved in the evaluation of the generic statement.
2. Alternatives of the group G the generic statement is about, $Alt(G)$, help to determine to what extent f is a *distinctive* feature for group G.
3. Alternatives, in the sense of causally background factors, influence our assessment of the extent to which (being a) G is causally relevant to f.

By taking the second and third type of alternatives very seriously, we will end up with an interpretation rule of which the majority rule is only a special case. The third set of alternatives also provides a straightforward link to experimental results showing that there is a close relation between judgements concerning generic sentences and general causal knowledge about the world (Murphy and Medin, 1985; Murphy, 2004). Again, we will argue here that a causal approach to generics should not be seen as a competitor to the statistical approach, but that both approaches are closely related (in contrast, for instance, to what is claimed in Cimpian et al., 2010b).

Our argumentation will proceed step-wise, starting with the first set of alternatives in section 2.1, continuing with the second set in section 2.2, and finally introducing the third notion of alternatives in section 2.3. So, in the first part of the paper each section will end with a new, extended version of the majority rule just introduced. In the second part of the paper we will zoom in on the second type of alternatives we have added and provide additional support for our claim that they play a role for the interpretation of generics. In section 3.1, we will connect our approach to BP generics to the analysis of conditioning in the psychology of learning. This leads to a last adaption of the approach to generics defended here, introduced in section 3.2. In section 4, we will present the results of two experiments testing this approach to BP generics.

2. THE DIFFERENT WAYS GENERICS DEPEND ON ALTERNATIVES

2.1. Alternatives to Determine the Probability Domain

The most straightforward way to link the truth conditions of generic sentences to statistical data is the majority rule introduced in section 1: to account for the truth of (1) we demand that the *majority* of birds fly. We already discussed in the introduction an example showing that such an account doesn't work in general. Examples like (3-a) and (3-b) make the same point. Again, these generics are acceptable, even though $P(f|G)$ seems to be less than half.

(3) a. Ducks lay eggs.
b. Goats produce milk.

However, these examples can be given a majority analysis after all, if we make an additional use of alternatives (cf. Cohen, 1999). The relevant alternatives for a generic of the form "*G*s are f" will be alternatives to the feature f, i.e., $Alt(f)$. For (3-a), for instance, we should take into account $Alt(lay\ eggs)$. Intuitively, $Alt(lay\ eggs)$ will consist of alternative ways of reproduction. Thus, $Alt(lay\ eggs) = \{lay\ eggs, give\ live\ birth\}$. Cohen (1999) proposes that the probability function relevant for the interpretation of the generics should now not range over *all* objects, but be restricted to the set of objects that satisfy at least one of the properties in $Alt(f)$, i.e., $\bigcup Alt(f)$. We end up with the following adaption of our stable majority rule.

Definition 2. *Truth conditions for generics with* $Alt(f)$ *alternatives. A generic sentence "Gs are f" is true in context c in case for the contextually salient set* $Alt(f)$ *of alternatives to f it holds that*

$$P(f|G \cap \bigcup Alt(f)) > \frac{1}{2}.$$

Because $\bigcup Alt(lay\ eggs) \approx Females$, a majority analysis could, or would, predict that (3-a) is true just in case a (stable) majority of *female* birds lay eggs[1].

Unfortunately, as already known by Cohen (1999), definition 2 won't do. There are various other examples where this application of alternatives won't save the majority rule. In general, a high conditional probability of f given G appears to be neither a *sufficient*, nor a *necessary* condition for the corresponding generic to be true. As for necessity, it is unclear how even the new Definition 2 could explain example (2) from the introduction, or example (4), which is very similar to (3-a) and (3-b).

[1] Although such an analysis seems natural, it is not the strategy that Cohen (1999) suggests to account for examples like (3-a) and (3-b). Instead, Cohen (1999) proposes that these type of sentences should be treated as relative readings, to be discussed in the next section.

(4) Ducks have colorful feathers.

The following type of examples, mostly due to Carlson (1977), have been used to show that a high conditional probability is not a sufficient condition either:

(5) a. *Chicken are female.
b. ?Chinese speak Mandarin.
c. ?People are over 3 years old.
d. ?Crocodiles die before they attain the age of 2 weeks.
e. ?Primary school teachers are female.
f. ?Bees are sexually sterile.

Although these generic sentences all seem false, or at least not (obviously) true, their corresponding conditional probabilities are high. In particular, although about 80% of all chicken are female, due to the fact that, for economic reasons, most farmers gas male chicks immediately after birth, the generic (5-a) seems false. In all these cases the amended majority rule proposed in Definition 2 is of no help. A similar point can also be made with the following two famous examples.

(6) a. ?Books are paperbacks.
b. ?Mammals are placental mammals.

Again,the approach fails, because most naturally,

$$\bigcup Alt(paperbacks) = \bigcup\{paperbacks, hard\text{-}covers\} \subseteq Books,$$

with the result that (6-a) is still falsely predicted to be true, if the majority of books are paperbacks. But there might be another way to go. Perhaps we can demand that for a generic of the form "*G*s are *f*" to be appropriate, it cannot be the case that $\bigcup Alt(f) \subseteq G$. This constraint would immediately rule out examples like (6-a) and (6-b) and some other weird generics like "Humans are autistic," which would be predicted to be inappropriate, instead of just false, simply because only humans can be autistic (or let us assume so). This constraint certainly helps with some of the counterexamples to sufficiency. But it is of little help when it comes to examples like (5-a). Additionally, so far we miss a rationale behind this constraint, though we will provide one in section 3.2.

In the following section, we will discuss the use of two more sets of alternatives in the definition of the truth conditions of generic sentences. The first set will be used to account for the examples that show that high conditional probability is not a necessary condition for the truth of a generic. The second set will be used to explain why it is not a sufficient condition either.

2.2. Subject Term-Alternatives and Relative Readings

Let's have a look at a different class of very famous examples. Much ink has been spilled on the following "Port-Royal" type of generics:

(7) a. Dutchmen are good sailors;
b. Bulgarians are good weightlifters.

Intuitively, the above sentences are appropriate, although only a small percentage of Dutchmen are good sailors and only few of all Bulgarians are good weightlifters. It is also not the case that limiting the domain of the probability function to $\bigcup Alt(good\ sailor)$ would make (7-a) true on a majority analysis after all, because naturally $\bigcup Alt(good\ sailor)$ could include also things like "soldiers," "(good) peasants," etc.. One can imagine several strategies to deal with such sentences[2]. For instance, one might propose that limiting the domain to $\bigcup Alt(good\ sailor)$ would still do: Because in a natural use of (7-a) the adjective 'good' typically is stressed, the set $Alt(good\ sailor)$ would typically be just $\{good\ sailors, moderate\ sailors, bad\ sailors\}$. Thus, $\bigcup Alt(good\ sailors) = Sailors$, meaning that the domain of the probability function would range only over sailors. It follows that (7-a) is predicted to be true on a majority analysis just in case most Dutch sailors are good sailors.

This solution, however, appears to be not particularly convincing. The reason is that although Bulgarian weightlifters are pretty successful at the olympics, it is questionable whether most Bulgarian weightlifters are good weightlifters. Similarly, it is questionable whether most Dutch sailors are (or were in the seventeenth century) good sailors. A much more natural solution seems to be to propose (perhaps with Nickel, 2012) that (7-a) is true just because the good Dutch sailors are good compared to good sailors in general *and* the moderate Dutch sailors are good compared to moderate sailors in general *and* the bad Dutch sailors are good compared to bad sailors in general. Interestingly, this reading is close to Cohen's (1999) analysis of sentences like (7-a) as *relative readings* of generics.

Cohen (1999) proposed that generics like (7-a)–(7-b) are true, because they should be interpreted differently than standard generics, namely in a *relative* way: (7-a) is true iff compared to relevant alternative people in the seventeenth century (Frenchmen, Spaniards, Englishmen, and people from the Germanic countries), *relatively many* Dutchmen are good sailors. Similarly for (7-b). In probabilistic terms this means that $P(f|G) > P(f)$—or better $P(f|G \cap \bigcup Alt(f)) > P(f|\bigcup Alt(f))$—should hold with "*G*" denoting the Dutchmen and "*f*" standing for "are good sailors.' "Making use of relative readings, we could also account for the fact that examples like (4) are, intuitively, true.

Cohen (1999) links the two readings of generic sentences to particular intonation patterns of the sentence used. If in the use of a generic sentence of the form "*G*s are *f*" it is the feature *f* that is stressed by intonation, the generic sentence will have a standard (stable) majority reading. But if (topical) stress is given to the

[2] According to one of them, what counts is not whether, for (7-a) for instance, the majority of Dutchmen *actually* are good sailors, but whether they *can* or *would* be good sailors if they tried. Although such a strategy might look appealing, the strategy seems to over-generate enormously: why, then, is an example like "Children are dangerous" not true, just because these children *can* be dangerous? According to another strategy, one might say that these sentences are actually false. But why, then, do so many people take them to be true? A major worry here is to determine what the data are: if (7-a)–(7-b) are generally taken to be true, what is it that makes the claim "correct" that these sentences are *in fact* false? It cannot be that this is so because it is predicted by the theory, because the theory itself is based on intuitions of the language users.

subject term "G," the relative reading follows. It is standardly assumed that topical stress indicates a contrast between that what is stressed, and the alternatives of the stressed item. The stress on G then indicates a contrast with denotations of other terms $G_1, \cdots, G_n$: compared to the alternatives of G, i.e., $G_1, \cdots, G_n$, many Gs have feature f. This suggests that the generic "Gs are f" is true in that case only if $\forall i: P(f|G) > P(f|G_i)$, or perhaps, only if $P(f|G) > P(f|\bigcup\{G_1, \cdots, G_n\})$[3]. If we assume that the "domain" of the probability function is $G \cup \bigcup\{G_1, \cdots G_n\}$ and that G is incompatible with all the G_i, the latter suggestion comes down to the requirement for "Gs are f" to be true that $P(f|G) > P(f|\neg G)$. Interestingly enough, it can be easily proved that $P(f|G) > P(f)$ if and only if $P(f|G) > P(f|\neg G)$, and thus that "Gs are f" is true on Cohen's relative reading exactly if $P(f|G) > P(f|\neg G)$. Hence, we can derive the relative meaning from a more general and independently motivated approach to the interpretation of focus.

Taking all that has been said about the relevance of alternatives for the meaning of generics into account, we end up with the following definitions of the truth conditions of generic sentences.

Definition 3. *Truth conditions for generics with* $Alt(f)$ *and G-alternatives.*
A generic sentence "Gs are f" is ambiguous between an **absolute** *and a* **relative reading**. *In its absolute reading the conditions of Definition 2 apply. In its relative reading the generic is true, in context c in case for a contextually salient set* $Alt(f)$ *of alternatives to f and a contextually salient set* $Alt(G)$ *of alternatives to G it holds that*

$$P(f|G \cap \bigcup Alt(f)) > P(f|\bigcup Alt(G) \cap \bigcup Alt(f)).$$

Suppose that a generic has a relative reading. In that case it is clear that high conditional probability is not a sufficient condition for the corresponding generic to be true. For instance, it might be that although $P(f|G)$ is high, still $P(f|G) < P(f|\neg G)$. Perhaps we could account for the falsity of the following sentences, by assuming that they receive a relative reading.

(8) a. *Chicken are female.
b. ?Chinese speak Mandarin.
c. ?People are over 3 years old.
d. ?Crocodiles die before they attain the age of 2 weeks.
e. ?Primary school teachers are female.
f. ?Bees are sexually sterile.

Although we think that it is quite natural that these sentences receive a relative reading, that won't help to predict all these sentences to be false: although it might explain why (8-c) is bad[4], (8-b), for instance, would obviously be true on its relative reading as well.

To account for these type of examples, Cohen (1999) and Cohen (2004) proposes a *homogeneity condition*. Rather than just demanding (for the absolute reading) that $P(f|G)$ is high[5]. Cohen demands that the conditional probability of f given a set of Gs should be high for each cell of a contextually determined salient partition $\{G_1, \ldots, G_n\}$ of G. Thus, each of $P(f|G_1) \cdots P(f|G_n)$ should be high. Although it is not usually thought of in that way, each cell G_i could, in fact, be thought of as an alternative. Concentrating on (8-f), for instance, a salient partition of bees into queens (female), workers (female) and drones (male) will correctly predict that (8-f) is false, because neither queens nor drones tend to be sterile. Cohen provides a similar explanation for other examples as well.

We think this proposal is promising, and we are sympathetic to this proposal because making use of the homogeneity condition fits well with our idea that generic sentences express inductive generalizations about unbounded sets (cf. section 4). Still, Leslie (2008) has persuasively argued that the condition of homogeneity not only explains away bad generics, but good ones as well. Why, for instance, is "Bees reproduce" true on Cohen's salient partition of bees?[6] More dramatically, consider (1) "Birds fly." This generic is predicted to be false on both readings, if the relevant partition is a bi-partitioning of birds into Penguins, on the one hand, and all the other types of birds, on the other. Why is this partition not the relevant one? Of course, Cohen could claim that this partition is not the salient one with respect to which the sentences should be interpreted, but then the question is, why not?

2.3. Alternative Causal Background Conditions

In van Rooij and Schulz (2019, 2020b), we have argued that many generics should be given a *causal* analysis. It is not the conditional probability that should be high in order for a generic of the form "Gs are f" to be true, it should rather be the case that having property G has a *significant causal impact* on also having feature f[7]. Intuitively, "Gs are f" is true on this analysis, if being a G, or having property G, is causally sufficient (with

[3] Forgetting for simplicity now about $Alt(f)$.

[4] By taking other large mammals as alternatives.

[5] In contrast to Cohen (1999) we will in this section interpret a conditional probability like $P(f|G)$ as ranging over open formulas. Thus, $P(f|G)$ really measures the amount of Gs that are also f. Cohen (1999) rightly observes that in this way the "unbounded" character of generics cannot be accounted for. We agree, but we will propose our own remedy to solve this problem.

[6] Leslie (2008) wonders how a proponent of a probabilistic account can explain why the generic "Bees reproduce" seems true, while "Bees are sterile" is false. The problem is that if "Bees are sterile" is (correctly) predicted to be false because it is not the case that the conditional probability $P(Sterile|Bee)$ is high for all types of bees, the generic "Bees reproduce" is for that reason (wrongly) predicted to be false as well, because members of at least one type of bee (the workers) don't (tend to) reproduce. We think that "Bees reproduce" is nevertheless true, because in many cases plurals like "bees" and "ants" are seen as mass nouns and have a *collective* interpretation due to the fact that these are very small insects that we most of the time don't individuate (cf. van Rooij and Schulz, 2020a). On such a (semi-) collective interpretation of "Bees reproduce," it doesn't have to be the case that all (minimal) subgroups of bees reproduce, it is only required that the whole group—or better, larger subsets of this group—does so. Notice that although in English, "bee" and "ant" are count nouns, their counterparts in languages such as Welsh (Stolz, 2001) and Dagaare (Grimm, 2009) are actually mass nouns. This suggests that it is at least natural to view bees and ants primarily as collections.

[7] To be sure, we don't think that all generics have such a causal interpretation, but we think that many of them have.

high probability) for also having feature f. The notion of "*causal impact*" is defined by Pearl (2000) in terms of intervention, making use of causal models. Fortunately, we can reformulate (or test) this notion without making use of interventions by making use of alternatives.

In causal models there exists a difference between the probability of C conditional on the observation of A and the probability of C conditional on making A true by intervention. The former is modeled by standard conditionalization, $P(C|A)$. The latter, however, is modeled by $P(C|do(A))$. Whereas, $P(C|A)$ has a purely evidential reading, $P(B|do(A))$ has a causal one. An appealing way to illustrate the difference between $P(C|A)$ and $P(C|do(A))$ is by making use of partitions (Skyrms, 1980; Pearl, 2000). According to standard probability theory, $P(C|A) = \sum_i [P(C|B_i \wedge A) \times P(B_i|A)]$, with $\{B_i\}$ any partition of the state space. Instead, $P(C|do(A)) = \sum_i [P(C|B_i \wedge A) \times P(B_i)]$, where the B_i are the maximally specific causally relevant background factors[8]. Notice that although in general $P(C|A) \neq P(C|do(A))$, they come to the same if A is probabilistically independent of the issue of which causal background factor in fact holds, i.e., if for all B_i, $P(B_i|A)$ is the same as $P(B_i)$.

In section 2.2, we have seen that according to Cohen (1999) "Gs are f" is true on its relative reading iff $P(f|G) - P(f) > 0$, which is equivalent with $P(f|G) - P(f|\neg G) > 0$ (where $\neg G$ stands for $\bigcup Alt(G)$). If we would say that "Gs are f" is true iff having property G has a *positive causal impact* on also having feature f, this comes down to demanding that $P(f|do(G)) - P(f|do(\neg G)) > 0$. This already shows that the relative reading is closely related to the causal reading of generics. In fact, Cohen's relative reading can be seen as a special case of our causal reading. To see this, notice that in terms of causal background factors, the condition $P(f|do(G)) - P(f|do(\neg G)) > 0$ reduces to $[\sum_i P(f|G \wedge B_i) \times P(B_i)] - [\sum_i P(f|\neg G \wedge B_i) \times P(B_i)] > 0$. If the issue $\{G, \neg G\}$ is independent of the issue which causal background factors in fact hold, this, in turn, comes down to $[\sum_i P(f|G \wedge B_i) \times P(G|B_i)] - [\sum_i P(f|\neg G \wedge B_i) \times P(\neg G|B_i)] > 0$, which reduces to Cohen's relative reading: $P(f|G) - P(f|\neg G) > 0$.

We have stated above that the causal impact of G should not just be positive, but should rather be *significant* in order for the generic to be true: the difference should be *significantly* above 0. Thus, we end up with the following causal analysis of generics:[9]

Definition 4. *The generic sentence "Gs are f" is true iff*
$$\sum_i [P(f|G \wedge B_i) \times P(B_i)] \gg \sum_i [P(f|\neg G \wedge B_i) \times P(B_i)],$$
where $\{B_i\}$ is a partition of maximally specific causally relevant background factors.

Notice that each causal background factor B_i of the partition $\{B_i\}$ can be thought of as an alternative, in a similar way as each cell G_i of the salient partition $\{G_1, \cdots G_n\}$ used in Cohen's homogeneity condition can. We don't know whether the causal background partition can replace Cohen's homogeneity condition, but if so, it would explain why the partition $\{Penguins, other\ birds\}$ is not a good partition with respect to which "Birds fly" must be interpreted, if we (with Skyrms, 1980) additionally demand that $\forall B_i : P(f|G \wedge B_i) \geq P(f|\neg G \wedge B_i)$. In any case, we think that a causal analysis, and thus our causal alternatives, can help to explain why some of (8-a)-(8-f) are false.

Take an example like (8-b). Obviously, a large population of Chinese speak Mandarin, so $P(M|C)$ is high, and much higher then $P(M|\neg C)$. But on our causal analysis, we must compare $P(M|C \wedge B_i)$ with $P(M|\neg C \wedge B_i)$ for the B_i that are causally relevant for whether or not somebody speaks Mandarin. Whether or not you live in China, or communicate a lot with people that live in China, seems a natural candidate. But when B_i stands for "living in China," the difference between $P(M|C \wedge B_i)$ and $P(M|\neg C \wedge B_i)$ doesn't seem to be that high. On the other hand, $P(B_i|C)$ is high (and $P(\neg B_i|C)$ is low) and very different from $P(B_i)$. Thus, there is a difference between the evidential impact, $P(M|C) - P(M|\neg C)$, on the one hand, and the causal impact, $P(M|C \wedge B_i) - P(M|\neg C \wedge B_i)$, on the other: whereas the former difference is high, the latter difference is (presumably) low. But that is enough to explain why (8-b) is false, if we assume that the generic has a causal interpretation.

Other examples can be explained (away) in similar ways. Consider for instance (8-e), "Primary school teachers are female." This sentence is predicted to be false on a causal interpretation, because there doesn't seem to be any B_i that is causally relevant for being female such that $P(F|PST \wedge B_i) - P(F|\neg PST \wedge B_i)$ is high, though being a primary school teacher is still evidentially relevant for the most natural partition $\{B_i\}$, i.e., the genetic makeup.

Before we conclude this excursion into causality, note that the analysis of generics we propose here combines a causal analysis of generics with a probabilistic approach. We want to highlight this because, as mentioned in the introduction, the shortcomings of the majority rule are sometimes interpreted as showing that a statistical approach of generics is doomed to fail (cf. Leslie, 2008; Cimpian et al., 2010b). However, this is fallacious reasoning. There are many more options that one can take when exploring statistically approaches than just the majority rule. And the observed connections between the truth conditions of generics and assumed causal dependencies can also be captured nicely with a statistical approach. We will come back to this point in the next section, when we discuss the relation of generics to associative learning.

Furthermore, notice that the approach proposed here can, for instance, also account nicely for some of the experimental data on the dependence of generics on causal world knowledge. Cimpian et al. (2010b) reports that generics based on biological features are judged true more often than generics based on more accidental features (having a broken leg, or having infected ears). The generics based on biological features were also assumed to imply a significantly higher probability of the feature in the group than generics based on accidental features. Such

[8]This is the way Pearl (2000) estimates $P(C|do(A))$ when no explicit intervention, or experiment, is possible. B is thought of as the confounding variable that should be controlled.

[9]In van Rooij and Schulz (2019, 2020b), a slightly different notion is used, the notion of 'probability of causal sufficiency'. One can show that under some natural conditions this comes down to $\frac{P(f|do(G)) - P(f|do(\neg G))}{1 - P(f|\neg G)}$ – which is basically the same as Cheng's (1997) notion of Gs 'causal power' to produce f. Although to determine the numeral value of causal power, the denominator is important, we will ignore this denominator in this paper.

generics would also have a hard time passing the truth conditions proposed in Definition 4.

3. GENERICS AS LEARNING GENERALIZATIONS

In this section we will focus on the second sense in which generics take alternatives into account: alternatives to the group G the generic claim is talking about. The alternative set $Alt(f)$ will be put aside for the moment. In the first subsection below we will show that the semantics proposed by Cohen for the relative reading is strongly related to how in Psychology associative learning is described. This leads to an interesting new perspective on the meaning of generic sentences: we should understand their meaning in terms of the conditions under which we would learn the expressed generalization. This would give a natural explanation for why theories of learning appear so relevant for the meaning of generic sentences.

However, in two important ways this perspective does not mesh well with the approach we finished section 2 with. First of all, learning is something that grows gradually with the experience of the learner. There is no clear cut-off point in contrast to what Definition 3 assumes for both readings of generics. Second, the results from learning motivate the relative, not the absolute reading of generics that Cohen postulates. These two considerations will lead us to formulate an alternative approach to generics in section 3.2. This is the approach that will then be tested in the final section of the paper (section 4).

3.1. Subject Term-Alternatives and Learning

In this section we argue that there is an important justification for assuming that generic sentences (also) have a relative reading, and thus that the subject alternatives $G_1, \cdots, G_n$ matter for the interpretation of a generic sentence. In section 2, we have stated that generic sentences express, by their very nature, useful generalizations. This suggests that there is a close relation between the truth conditions of generic sentences, on the one hand, and the way we *learn* generalizations, on the other. Much psychological research on learning was done before the cognitive revolution in psychology, in classical conditioning. In classical conditioning, what is learned is an association between a cue and an outcome. The cue, c, such as the sound of a bell, or a tuning fork, can become associated with an outcome, o, which can be thought of either as something like the taste of food, or a shock, or an unlearned reflex response to that, like salivation, or high blood pressure indicating fear.

What is the expectation that the $n + 1$th cue c will be accompanied with outcome o? The perhaps most natural idea would be that it is just the times that cue c was accompanied with outcome o divided by the times that cue c was given at all. If we say that $V_i(o|c) = 1$ if at the ith exposure cue c is accompanied with outcome o, and that $V_i(o|c) = 0$ if at the ith exposure cue c is not accompanied with outcome o, the expectation according to this natural idea that the $n + 1$th cue c will be accompanied with outcome o, i.e., $P^*_{n+1}(o|c)$, can be stated as follows:

$$(RF) \qquad P^*_{n+1}(o|c) = \frac{V_1(o|c) + \cdots + V_n(o|c)}{n} = \frac{1}{n}\sum_{i=1}^{n} V_i(o|c)$$

It is well-known, however, that for the calculation of $P^*_{n+1}(o|c)$ it is not needed to maintain a record of all cases where cue c was accompanied with outcome o. One can calculate $P^*_{n+1}(o|c)$ incrementally as well, by constantly changing the expectations:

$$P^*_{n+1}(o|c) = \frac{1}{n}\sum_{i=1}^{n} V_i(o|c) = P^*_n(o|c) + \frac{1}{n}\Big(V_n(o|c) - P^*_n(o|c)\Big)$$

It turns out that the form of this incremental learning rule is very common. It is known as learning by *expected error minimization* and is used in almost all modern methods of learning.

Although it is natural to think that the expectation of outcome o for the $n + 1$th cue c will be $P^*_{n+1}(o|c) = \frac{1}{n}\sum_{i=1}^{n} V_i(o|c)$, this is not what is found experimentally, at least for animal learning. For animal learning, Rescorla (1968) observed that rats learn a tone (cue/cause)-shock (outcome) association if the frequency of shocks immediately after the tone is higher than the frequency of shocks undergone otherwise. This holds, even if in the minority of cases a shock actually follows the tone. Gluck and Bower (1988) and others show that humans learn associations between the representations of certain cues (properties or features) and outcome (typically another property or a category prediction) in a very similar way. Thus, we associate outcome o with cue c, not so much if $P(o|c)$ is high, but rather if $\Delta P^o_c = P(o|c) - P(o|\neg c)$ is high, where ΔP^o_c is known as the *contingency* of o on c. How can this be explained? Rescorla and Wagner (1972) show that this can be explained by an error–based learning rule very similar to the one above. The only thing that really changes is that this time the learning rule is also *competition-based*. The idea is that a cue can also be taken as a *combination* of separate cues: if c_1 and c_2 are cues, c_1c_2 is taken to be a cue as well, and they all could be accompanied with the same outcomes. According to Rescorla and Wagner (1972), we should keep track of expectations, or associations, for cue-action pairs for all primitive cues, i.e., c_1 and c_2. For the calculation of $E^*_{n+1}(o|c_1)$ after the nth trial, however, we should also look at $E^*_{n+1}(o|c_2)$ in case the actual cue at the nth trial is the combined cue c_1c_2. The famous Rescorla-Wagner learning rule (RW) for each primitive cue c_i is stated as follows:

$$(RW) \qquad E^*_{n+1}(o|c_i) = E^*_n(o|c_i) + \lambda\Big(V_n(o|c^*_i) - \sum_j E^*_n(o|c_j)\Big)$$

Here, $E^*_{n+1}(o|c_i)$ is the agent's expectation after n observations that the $n + 1$th primitive cue c_i has outcome o, where λ is a learning rate (typically very small) and where $V_n(o|c^*_i)$ measures the magnitude of the reinforcement at the nth trial where cue c_i was involved[10]. Although $E^*_{n+1}(o|c)$ converges to the

[10]Take c_i to be c_1. Then it could be that the actual cue was c_1c_2 and that $V_n(o|c_1c_2) = 1$, although $V_n(o|c_1)$ would be 0.

actual conditional probability (or relative frequency) under some conditions, Cheng's (1997) shows that under most conditions $E^*_{n+1}(o|c)$ yields, instead, $\Delta P^o_c = P(o|c) - P(o|\neg c)$ in the long run (see also Danks, 2003). Thus, in those cases expectations, or associations, as generated by rule (RW) do not really measure probabilities, but contingency, instead[11] We have noted already that $\Delta P^o_c = P(o|c) - P(o|\neg c) > 0$ if and only if $P(o|c) > P(o)$, i.e., the measure Cohen (1999) used to account for relative readings of generics. Interestingly, Yuille (2006) shows there exists a learning rule very similar to (RW) that converges to Cheng's (1997) notion of causal power, which is closely related with the notion of "causal impact" as discussed in section 2.3. Thus, not only Cohen's relative reading can be motivated through learning, the causal analysis of generics sketched in section 2.3 can be given a learning-theoretic motivation as well.

3.2. A New Proposal

Based on the discussion in the last section, we propose that the truth, or assertability, of generic sentences should be stated in terms of the conditions needed to learn the expressed generalization. More concretely, we want to propose (but see also van Rooij and Schulz, 2020a), that the measures used in the above discussed literature on learning can also be used to measure the assertability of generic sentences. To have a concrete measure to work with we take contingency, instead of the more general notion of causal impact. If for simplicity we also ignore the alternative set $Alt(f)$, this gives the following proposal for the assertability of generic sentences.

Definition 5. *The assertability of a generic sentence "Gs are f" is given by the formula*

$$\textit{Assertability of 'Gs are f'} \quad = \quad P(f|G) - P(f|\bigcup Alt(G)).$$

We propose here that distinctiveness is at the heart of the meaning of generic sentences. Tessler and Goodman (2019) came up with a very similar proposal. Our motivation, however, is different: we propose Definition 5 because of the close connection between the meaning of generic sentences and how we learn (causal) generalizations. Definition 5 differs from the interpretation rule we ended up with in Definition 3 in that it replaces truth conditions for generics with degrees of assertability. We think that this is a step that we have to take. From a theoretical point the use of cut-off points seems necessary to allow for a truth-conditional approach to generics. This strategy to translate grades into a binary system occurs in semantics and philosophy of language at various points (vagueness, conditionals, etc.), but it is also known to be very problematic: a vague predicate is vague exactly because it does *not* seem to have a clear cutoff point. It doesn't seem to be convincing at all that we switch our ratings of assertability of sentences completely based on small differences in the frequencies that we observe. For similar reasons, and because of the link we want to make to associative learning, we propose here that at least the assertability of generics is a matter of degree. We don't want to engage in a discussion of what that would mean for truth conditionals semantics in general here. This will be left for future work.

Another important difference with Definition 3 is that the relative reading introduced there[12] now becomes the base case for generic sentences. As noted above, in this respect we agree with the closely related proposal of Tessler and Goodman (2019)[13]. One might wonder what happened to the absolute reading that Definition 3 talked about? Does it disappear in the new approach? Not at all. We want to propose that the absolute reading now re-emerges as a special case of the interpretation rule given above. In case there are no salient alternatives to the group G, the factor in the equation that is due to these alternatives disappears and the assertability of generic sentences is entirely measured in terms of the conditional probability of f given G[14,15].

As noted above, our proposal in this section is a special case of the causal analysis proposed in section 2.3. However, for the rest of the paper we will work with the somewhat simplified approach stated in Definition 5. This approach can account for the same examples that the proposal in section 2.2 can deal with. But we also get something extra. Taking a relative reading as the underlying and general meaning of generic sentences allows us, for instance, to account for the fact that the generic (9) seems false, or at least inappropriate in most situations. There is hardly any set of alternatives that would explain why there is anything special about Germans as far as right-handedness is concerned. On the other hand, talking about Germans seems to evoke very naturally comparison to other nationalities. So, it is hard to imagine a context in which such alternatives wouldn't be considered at all. But if such alternatives are salient, then the proposal above would predict the generic (9) to be not assertable.

(9) ?Germans are right-handed.

The proposal also provides a way to understand the constraint $\bigcup Alt(f) \not\subseteq G$ we discussed to account for the oddness of examples like (10-a) and (10-b).

(10) a. ?Books are paperbacks.
b. ?Mammals are placental mammals.

According to this constraint these generics are odd, because the relevant feature (being a paperback) only applies to the targeted group (books). Assuming that generics are about

[11] We take it to be very natural, however, that people take the associations, to be the conditional likelihood. In fact, according to, e.g., Newel et al. (2007), we can explain many of the problematic probability judgements as found in, e.g., Tversky and Kahneman (1974) by the assumption that people confuse probabilities with associations as established via associative learning mechanisms like (RW). See van Rooij and Schulz (2020a) for a use of this idea for the analysis of generics.

[12] According to Cohen (1999) this is the less important reading of generics.

[13] Although they don't base or motivate their proposed analysis on learning-theoretic grounds.

[14] If one assumes that $P(f|X) = 0$, if $X = \emptyset$, this straightforwardly follows from Definition 5. For a more principled motivation, see Cheng's (1997) and van Rooij and Schulz (2019).

[15] Notice that this still doesn't mean that the assertability of generics does come down to the majority rule in this case. We keep the claim that assertability comes in grades (the grade is given by the conditional probability $P(f|G)$) and don't assume a cut-off point of 0.5.

distinguishing the group with the feature, together with well-established pragmatic constraints allows us now to make sense of this constraint. The pragmatic assumption we need is the Gricean rule that the sentence uttered needs to be informative. Notice that in the cases discussed here the fact that all objects with property f are part of group G is *a priory* knowledge: it is part of the meaning of these words. In other words, without observation you already know that all f are G. Therefore, the claim made by the generic according to Definition 5 that f is distinctive for G is not informative and, thus, out for pragmatic reasons.

4. EMPIRICAL RESULTS ON THE ROLE OF *G*-ALTERNATIVES

In the previous sections we have argued in favor of the claim that alternatives are relevant for the interpretation of a generic sentence of the form "*G*s are *f*" for several reasons: (i) alternatives to f are relevant to restrict the domain of the probability function; (ii) alternatives to the subject term G are relevant in case the generic has a relative, or contrastive reading, and (iii) alternative causal background factors influence our assessment of the extent to which (being a) G is causally relevant to f. Moreover, we have argued that alternatives to the subject term G are important in any case to learn the (inductive) generalization. We have motivated the importance of these sets of alternatives by looking at core examples in the literature. For the second set of alternatives we also provided independent evidence coming from the field of psychology of learning. In this section we will present the results of three empirical studies on the relevance of *G*-alternatives for the interpretation of generics. Ultimately, this should be done for the other sets of alternatives as well, but this will have to wait for future work.

4.1. The Hypotheses That We Will Test

The central goal of this part of our research was to empirically test whether alternatives to the subject term G do indeed affect the assertability of a generic sentence. Specifically, we hypothesize that the probability with which the alternatives carry the relevant feature f affects the assertability of the generic. This conforms with the account for generic sentences that we ended up with in section 3. According to this approach a generic *Gs are f* is the more assertable, the more distinctive the feature f is for the group G. The probability of f given G should be high *relative to* the probability of f given the salient alternatives to G[16].

Hypothesis 1. *The assertability of a generic sentence "Gs are f" depends on the conditional probability of the feature f given salient alternatives G' of G.*

To test this hypothesis, we manipulate $P(f|G')$ and see whether we can observe an effect on the assertability of the generic. Depending on whether or not this hypothesis is supported by the data, we can then test different approaches to the meaning of generic sentences that explain the result. For instance, if the observed assertability is in line with Hypothesis 1, then we can evaluate the particular rule that we formulated in Definition 5 for the assertability of generic sentences. In other words, we can test whether contingency is a good predictor for the assertability of generic sentences.

Hypothesis 2. *The assertability of a generic sentence "Gs are f" is given by the formula*

$$Assertability\ of\ 'Gs\ are\ f' = P(f|G) - P(f|\bigcup Alt(G)).$$

In the following, we will present the results of two experiments testing the hypotheses formulated above. We were looking for a setup that allowed us to probe the intuitions of people concerning generics about a group of objects for which they do not have any prior knowledge. This will allow us to ensure that participants do not have prior beliefs about features typical for the objects they will see. A second objective was to control the *G*-alternatives that the interpreters were considering. This is the factor that we will manipulate in order to see whether it influences the assertability of the generic sentence[17].

We presented participants with a picture-sentence verification task similar to that used in Bordalo et al. (2016). The participants saw pictures with samples of fictive insect species from two Galapagos islands, Genovesa, and Marchena (see **Figure 1**)[18]. Their task was to assess whether animals from one of the islands, Genovesa, could be described with a given sentence. All sentences were generics stating that the species from Genovesa—our target group G—has a particular feature having to do with their coloring—our target feature f. We controlled the conditional probabilities $P(f|G)$ that the participants of the studies assigned by manipulating how many of the animals G in the sample form Genovesa showed the particular coloring pattern f. The second sample from Marchena served as contextually salient alternative. By manipulating the frequency of insects with the relevant feature

[16]We want to emphasize that the feature that we are looking for here: the relevance of distinctiveness for the meaning of generic sentences, is in itself not distinctive for the particular approach we are defending here. A similar prediction is also made by approaches to generics that take them to be at the core assertions about kinds and thereby link them to how we represent kinds (Leslie, 2008). As, for instance, argued in Cimpian et al. (2010b) such an approach predicts that "…features that are privileged in our concepts may be more acceptable than generic predications of features that are not, all other things being equal." (Cimpian et al., 2010b, p. 1,456). Thus, "the more striking, appalling, or otherwise gripping we find the property predicated in the generic, the more tolerant the generic is to exceptions" (Leslie, 2008, p. 15). Distinctiveness is taking to be one of the characteristics that makes a property more gripping. We come to the same predictions about the relevance of distinctiveness, but via a different route. In our case it is linked to the learnability of the expressed causal dependence.

[17]This is not the first time the hypothesis that distinctiveness matter is empirically tested, see in particular (Cimpian et al., 2010b). Also in this paper novel categories are used to test the impact of contrastiveness on generics. They, however, ask people to judge the truth of a generic sentence, while we focus on assertability. Furthermore, they present the prevalence of the relevant feature in the target group verbally, while we provide this information graphically. We think that this is a more natural setting given the connection we want to make to learnability (though also our paradigm has its limitations, as we will discuss later). Furthermore, our experimental paradigm makes a much more fine-grained and precise manipulation of distinctiveness possible.

[18]The names of the islands are real. The participants were also shown a map of the Galapagos islands with the location of the islands. We chose animals instead of, for instance, manipulating the clothing of people, because the coloring of animals would not be perceived as an accidental feature of the observed individuals.

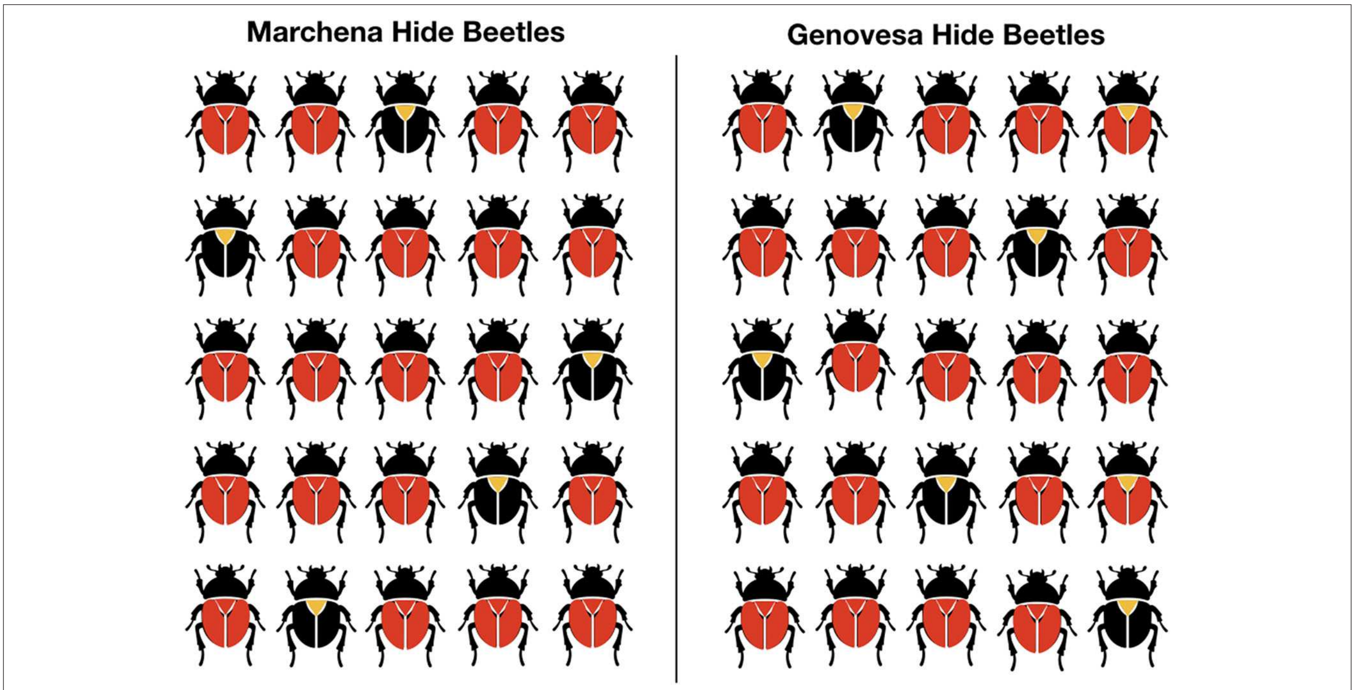


FIGURE 1 | Sample picture in the non-contrastive condition with beetles.

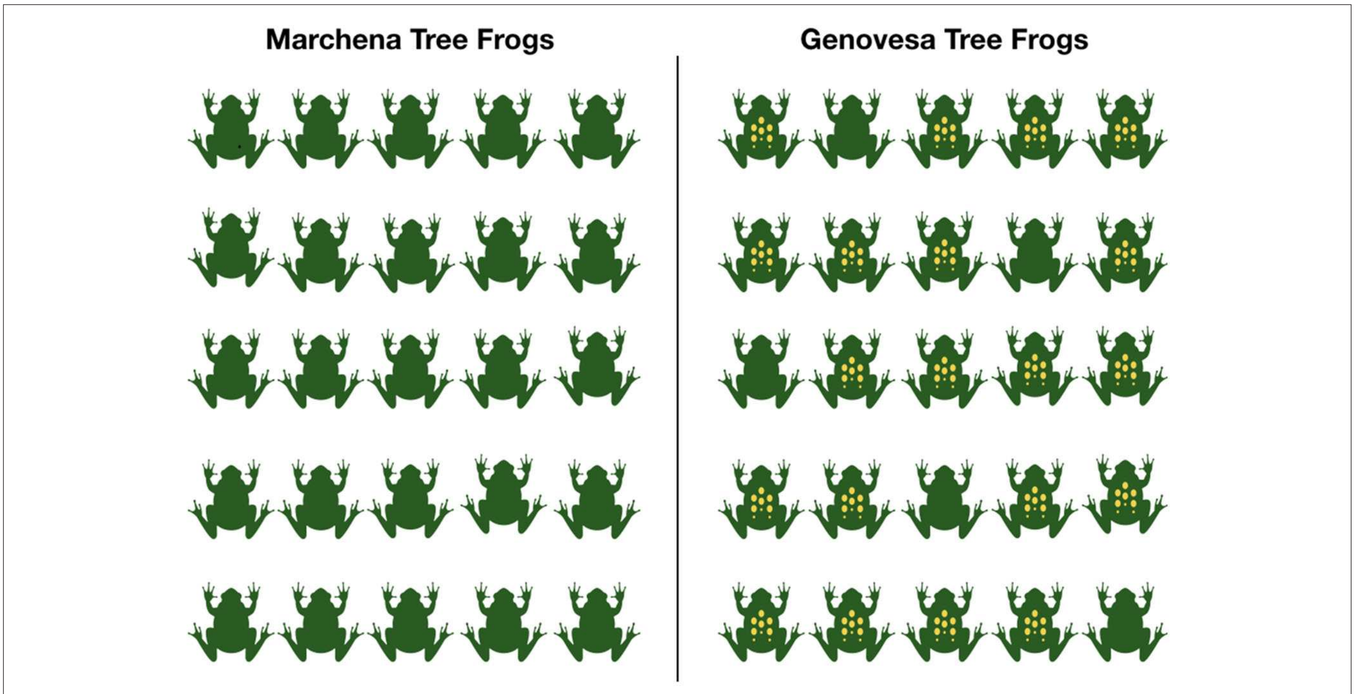


FIGURE 2 | Sample picture in the contrastive condition with frogs.

in this group we controlled $P(f|\bigcup Alt(G))$, from now abbreviated by $P(f|Alt(G))$.

We presented pictures in two conditions. In the non-contrastive condition an equal number of insects (80%) in both samples had the relevant feature f (see **Figure 1**). Thus, in this case $P(f|G) = P(f|Alt(G))$. In the contrastive condition, none of the insects in the sample from Marchena (the salient alternative) had the feature, while 80% of the insects from Geneva (the

FIGURE 3 | Example question from the study.

target G) had the feature f (see **Figure 2**). In other words, in this condition $P(f|G) = 0.8$ and $P(f|Alt(G)) = 0$. Based on Hypothesis 1, we expect that the strong difference of $P(f|Alt(G))$ between both conditions should have a significant effect on the assertability of the generic sentences. Hypothesis 2 predicts that the judgments of assertability people give for the generics should correspond to the contingency or the relative difference of feature f given group G.

4.2. Study 1

In the first study we used a within-subjects design. All participants gave an assertability score to one sentence in the contrastive condition, one in the non-contrastive condition and one filler sentence. Each question was presented with a different animal species (spiders, frogs, or bugs). Below the two samples, a generic sentence was given that always described the species from Genovesa. The participants were asked to judge on a scale from 0 to 5 whether the generic sentence was assertable given the provided data (e.g., *"Can you say the following to describe Tree Frogs from Genovesa?"*, see also **Figure 3**). They gave a response by dragging a slider as depicted in **Figure 3**. They could adjust their response with a accuracy of two decimals, so they experienced the scale as continuous.

Based on Hypothesis 1, we expected a significant difference in the judgments of assertability for both conditions. Hypothesis 2 claims that the judgments of assertability people give for the generics should correspond to the contingency of feature f given group G. In terms of proportions this measure predicts that the assertability of a generic should increase if feature f becomes more distinctive for the group G. Applied to the two conditions distinguished here we would expect that the generic is significantly more assertable in case of the contrastive condition than in the non-contrastive condition. The measure of contingency also makes precise numerical predictions for the assertability of generics. However, these predictions need to be translated into the scale presented to the participants in the study, because the range of the contingency function does not match the scale presented to the participants of the study: the contingency function ranges between -1 and 1, whereas the scale the participants saw let them grade the assertability of the sentences between 0 and 5. We used a linear transformation to map their responses directly onto the range $[-1, 1]$ of the contingency function. Thus, 0 on the scale corresponds to a contingency of -1, 2.5 to a contingency of 0, and 5 to a contingency of 1. If we apply this linear transformation to the conditions that the participants of our study saw, Hypothesis 2 predicts that in the non-contrastive condition the contingency of the generic is 0, thus the participants should move the slide to around 2.5 on the given scale. In the contrastive case the contingency is $P(f|G) - P(f|Alt(G)) = 0.8 - 0 = 0.8$. This corresponds to the value 4.5 on the scale the participants saw. Given that there will be variation in how participants interpret the scale, we did not expect exactly the values predicted by the measure of contingency. However, the general proportional prediction described above should be visible in the data.

4.2.1. Method

4.2.1.1. Materials and procedure

We used pictures of three different animal species (Tree Frogs, Hide Beetles, Jumping Spiders). For each species we designed a picture in the contrastive and in the non-contrastive condition. All the pictures contained two samples, one with 25 animals of the species from Marchena, one with 25 animals from the species from Genovesa. For each species we had one corresponding generic sentence: "Hide Beetles from Genovesa have red wings," "Tree Frogs from Genovesa have yellow dots," "Jumping Spiders from Genovesa have green backs."

The participants saw each animal species once, one in the contrastive condition, one in the non-contrastive condition and a third species as a filler. This resulted in three experimental trials per participant. In the filler condition, participants saw a generic that claimed the group to have a feature that none of the animals had (for instance, it could be the picture on **Figure 1** with the generic "Hide Beetles from Genovesa have green wings") and, therefore, this sentence was clearly not assertable. The filler condition was used to control whether participants completed the study in good faith: we excluded participants who gave a score above 1.5 in the filler condition as they likely did not pay attention in the other conditions either. The order in which the contrastive and the non-contrastive condition were shown was randomized. The filler always occurred last.

The study was implemented in Qualtrics. Participants started by reading the informed consent text and agreeing to taking part.

They then read the instructions. Average time spent on the task was 143 s.

4.2.1.2. Participants

Participants were recruited via Prolific.ac, an online platform aimed at connecting researchers and participants willing to fill in surveys and questionnaires in exchange for compensation for their time (Palan and Schitter, 2018). We recruited native English speakers (British and American English) who reported no vision impairments[19]. Eighty-two participants completed the task. Three participants were excluded: two because they did not give a response in one of the experimental items, one because they gave a score of 1.5 or above on the filler item. Thus, 79 responses were included in the analyses reported below.

Due to a mistake in the set up of the experiment, the participants were not forced to answer the filler questions. We therefore ended up with 27 participants who gave no response to the filler conditions. However, the slider was always at 0 by default, so these participants most likely simply agreed with the score 0 and therefore pressed "respond" without moving the slider. For this reason, we still included these participants in the analyses[20].

4.2.2. Results

The mean score given by the included participants in the filler condition was 0.04 (SD 0.16); the mean score in the contrastive condition was 3.51 (SD 1.06); and, finally, the mean score in the non-contrastive condition was 2.88 (SD 1.50). We performed a Bayesian paired samples *t*-test to test for the strength of evidence in favor of the null hypothesis (no difference between conditions) as opposed to the hypothesis that the score given by participants should be higher for contrastive than for non-contrastive condition using JASP software (JASP Team, 2018) with default priors. This analysis resulted in $BF_{10} = 104$, meaning that the data was 104 times more likely under our hypothesis than under the null hypothesis. Thus, the first study does lend support to Hypothesis 1 claiming that alternatives to *G* do affect the assertability of a generic sentence and the general prediction of Hypothesis 2 about the tendency of this dependency: comparing situations in which a feature is distinctive vs. ones where it is not distinctive for a group, the generic has a higher assertability in the situation in which the feature is distinctive.

In order to approximately evaluate compatibility of the observed scores with the predicted scores based on the Hypothesis 2, we investigated the 95% confidence interval (CI) around the mean in each condition, assuming a normal distribution. Note that the correct interpretation of 95% CI is that if we conducted our study multiple times with different participants and calculated a corresponding 95% CI for each group of participants, we would expect 95% of these confidence intervals to contain the true mean of the whole population. Thus, we expect that in 5% of the cases the confidence interval will not contain the true mean of the whole population. So it is possible that in our particular sample the CI does not contain this true mean. Note also that the assumption of normal distribution here is lenient. Given these considerations, the confidence intervals can give us only a rough idea of where the true value of the corresponding score in the population would lie.

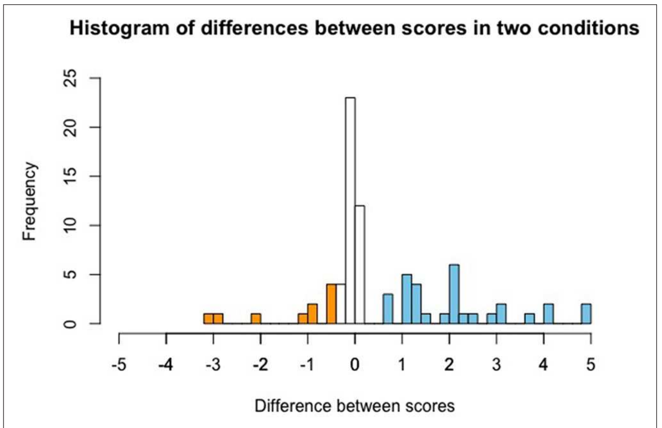


FIGURE 4 | Histogram of differences in scores between conditions: contrastive condition minus non-contrastive condition. Differences below −0.5 are marked in orange color, differences above 0.5 are marked in blue color. Orange bars thus indicate participants who gave a higher score in the non-contrastive condition, non-colored bars indicate participants who gave a similar score in both conditions, and blue bars indicate participants who gave a higher score in the contrastive condition.

We expected a mean score 4.5 in the contrastive condition, but observed 3.51 with 95% CI [3.27, 3.74] which does not include the expected score. For the non-contrastive condition, we expected a mean score 2.5, but observed 2.88 with 95% CI [2.54, 3.21] which again does not include the expected score, but does come close. Overall, while the scores come close to the expected ones, we cannot conclusively say that the observed values support our second hypothesis (but see the issues raised below in the *Interim Discussion* regarding the potential caveats of our approach).

Figure 4 depicts the difference between given scores in the contrastive and non-contrastive conditions for each participant (specifically, displayed is score in contrastive condition minus score in non-contrastive condition). We can see that not all participants uniformly gave higher scores to the contrastive as compared to the non-contrastive condition. In fact, there was a sizable proportion of participants who gave approximately the same score in the two conditions, and even a small group that gave the non-contrastive condition a higher score than the contrastive condition. Thus, we seem to be observing different behavior patterns by different participants. We will come back to this in section 4.4.

4.3. Study 2

The results of Study 1 supported the hypothesis that the score given by participants to assertability of a generic sentence will differ for the case with an alternative present and the case with no alternative present. The generic *"Gs are f"* becomes in general more assertable in case the discussed feature *f* is distinctive for the group *G*. The results also partially support Hypothesis 2:

[19]Since the material involved colors, the participants were required to have normal vision of colors.

[20]Excluding these participants did not make a difference to the results reported here.

in the non-contrastive condition the generic was judged to be in between assertable and non-assertable. In the contrastive condition the generic was on average rated to be assertable, though not to the degree predicted by the contingency measure. In order to replicate the original finding, we administered the same task in a between-participant set-up: each participant saw only one of the two conditions (contrastive or non-contrastive) plus a filler item.

4.3.1. Method

4.3.1.1. Materials and procedure

The materials used in this study were the same as in Study 1 except this time the participants saw only either contrastive or non-contrastive condition and a filler trial (2 trials in total). Average time spent on the task was 128 s.

4.3.1.2. Participants

Participants were recruited via Prolific.ac with the same eligibility criteria. One hundred eighty-two participants completed the task. Three participants were excluded from the analysis because of a missing response to one of the items. Further seven participants were excluded because of giving a score above 1.5 in the filler question. That left 172 participants for further analyses.

4.3.2. Results

The mean score given by the included participants in the filler condition was 0.07 (SD 0.23), in the contrastive condition 3.49 (SD 1.29; 95% CI [3.29, 3.68]), and in the non-contrastive condition 3.06 (SD 1.37; 95% CI [2.85, 3.26]). We performed a one-sided Bayesian independent samples *t*-test to test for the strength of evidence in favor of the null hypothesis (no difference between conditions) as opposed to the hypothesis that the score in the contrastive condition is higher than the score in the non-contrastive condition using JASP software with default priors. We obtained $BF_{10} = 2.5$, meaning that the data was 2.5 times more likely under the alternative hypothesis than under the null hypothesis. While this is not particularly strong evidence in favor of the alternative hypothesis, the data does show the same pattern as observed in Study 1. The diminished difference between conditions is likely due to that in Study 1, having two cases to compare, the participants noticed that the second set of objects changed (i.e., animals from Marchena), and this in turn strengthened the perceived contrast.

4.4. Interim Discussion

The results of both studies were in line with our Hypothesis 1: the probability of the feature f given a contextual salient alternative did affect the assertability of a generic sentence *"Gs are f"*. We also saw the direction of the dependence predicted by our theory supported: if $P(f|G)$ is substantially larger than $P(f|Alt(G))$ then the assertability of the generic sentence is higher than in case there is no difference between both probabilities. We did not see the exact assertability scores that the theory predicts (Hypothesis 2). In the non-contrastive condition, the theory predicts an assertability of 2.5, while in Study 1 the average assertability in this condition was 2.88 and in the Study 2 3.06 with 95% confidence intervals around mean not including the expected value in either case. In the contrastive condition, we predicted an assertability of 4.5 and observed an average of 3.51 in Study 1 and 3, 49 in Study 2, again with the 95% confidence intervals around the mean not including the expected value.

Contrary to our expectation, the participants were not uniform in the scores they were giving—we observed large differences between participants' behavior, so in fact it does not make much sense to look at the overall means as we set out when we started this project[21]. However, this observation does not necessarily contradict the theory tested here. The predictions made by contingency as measure of the assertability of generic sentences depends on which alternatives to G the interpreter considers. We assumed that the setup of the study would lead the participants to consider the sample from Marchena as alternative to the sample from Genovesa that the generic talked about. The theory predictions outlined above are only valid if the participants took the alternative into account. However, we cannot be sure that the participants really did take the sample from Marchena to be a relevant G alternative. If they did not take any alternatives to the target group into account, the theory predicts the assertability of the generic sentence to be equal to the conditional probability $P(f|G)$. Consequently, the assertabiility value assigned by the participants would be 4.

To explore this possible interpretation of the data, we separated the participants of Study 1[22] into three groups: those that assigned the same assertability rating to the generics in both conditions (difference between scores in the two conditions <0.5[23]), those that judged the generic in the contrastive condition to be at least 0.5 points more assertable and those who considered the generic at least 0.5 points less assertable. 51% ($N = 40$) of the participants in the first study did not give a substantially different score in the two conditions, while 38% ($N = 30$) considered the generic in the contrastive condition more assertable than in the non-contrastive condition and 11% ($N = 9$) of the participants took the generic to be less assertable. We then looked at the scores given by participants in the first two groups[24]. If Hypothesis 2 is correct but only participants in the group that gave a higher score to the contrastive condition took the sample from Marchena as an alternative to the sample from Genovesa, these participants should have given the scores predicted by Hypothesis 2 whereas the participants in the group that did not take into account the sample from Marchena should have given score 4 in both conditions (as discussed above). This was not the case. In the group of participants that gave a higher score in the contrastive condition than to the non-contrastive

[21]Note that we report the mean values and statistics with the whole group despite this since we committed to an analysis plan before we collected data.

[22]This was not possible for the second study since we used a between-participants setup in that case.

[23]This is an arbitrary threshold that we chose. We assumed that a difference of 0.5 could arise from the participants trying to drag the slider to the same point on the scale, whereas larger differences would necessarily arise from intentional positioning of the slider at different points of the slider.

[24]We will not discuss the participants in the third group which gave the non-contrastive condition a higher score than the contrastive condition further as we do not know why they behaved like that. They could have not understood the instructions or they could have changed their interpretation of the target sentence halfway through the experiment.

position, the average assertability in the contrastive condition was 3.86 (SD 0.79; 95% CI [3.57, 4.14]) whereas the average assertability in the non-constrastive condition was 1.72 (SD 1.22; 95% CI [1.28, 2.15]). Thus, even in this subgroup of participants, the scores come close to the ones predicted by the theory, but we do not observe the exact values predicted by Hypothesis 2. The group that did not see a difference gave a mean score 3.35 (SD 1.18; 95% CI[2.98, 3.71]) in the contrastive and a mean score 3.4 (SD 1.22; 95% CI [3.02, 3.77]) in the non-contrastive condition.

There are a couple of remarks we want to add about the discrepancies between the assertability values predicted by the theory and the data obtained in the study. First of all, it is difficult to say how exactly the participants interpreted the scale that we asked them to use to indicate the assertability of the generic sentences they saw. We tried to avoid the ambiguity by labeling the extremes of the scale verbally as "not at all" and "certainly," but cannot be sure what the participants did in case they were not sure about assertability of the sentence (when it is neither assertable nor non-assertable).

Depending on how the participants interpreted the scale, there might be also an issue with the way we interpreted the numerical values that our theory predicts (Definition 5). The range of the contingency function is the interval $[-1, 1]$. We took this to mean that -1 corresponds to a completely unassertable sentence, 1 to a sentence that is completely assertable and 0 describing the turning point from not assertable to assertable. This is how we translated the values of the contingency function to the scale that we presented to the participants of both studies. To some extend this is also supported by the data. The obviously wrong filler items got average assertability judgments that were very close to 0. However, there is no guarantee that even if the assertability of generic sentences can be described in terms of contingency, as we proposed, the values are interpreted in the linear manner that we assumed. Maybe a 0 for contingency already means that we wouldn't accept the sentence. To avoid such issues, we could show the participants a scale with numerical values from -1 to 1 instead 0 to 5 as we did here and see whether this affects their assertability judgments for the same set of test data. This will need to be taken up in the follow-up research.

To sum up, in general the results support the theory proposed here, though we did not see the exact scores that we expected. As discussed above, this could be because we did not transform the values from the theory to the scale seen by participants correctly. For this, more research in the future is necessary. What we can assess is in how far the theory explains the general tendencies in the data that we gathered, and in this respect the results are encouraging.

4.5. Study 3

The main goal of this final study was to test a different aspect of the theory developed in section 3.2. We repeat here for reasons of convenience Hypthesis 2, which contains the heart of the proposal.

Hypothesis 2. *The assertability of a generic sentence "Gs are f" is given by the formula*

$$\textit{Assertability of 'Gs are f'} = P(f|G) - P(f|Alt(G)).$$

So far, we have focussed on testing whether we can observe the predicted effects of manipulating the second argument of the measure of assertability. We saw that indeed $P(f|Alt(G))$ does affect the assertability of generic sentences and also that the kind of influence predicted (assertability goes up if $P(f|Alt(G))$ goes down) can be observed. In this study, we focused on the first part of the measure: $P(f|G)$. Manipulation of this factor should, according to our theory, also have an effect on the assertability of a generic. Roughly put, increasing this variable should have a positive effect on the assertability ratings.

As a side question, we also wanted to test with this study whether another new aspect of our proposal can be supported by the data. As discussed in section 3.2, the approach introduced in Definition 5 also differs from the one described in Definition 3 in measuring the assertability of generics in degrees instead of proposing cut-off points that define the limit between being or not being assertable. For instance, if alternatives do not play a role, then Hypothesis 2 predicts a steady linear increase in the assertability of the generic with growing $P(f|G)$. In some sense, the data of the first two studies already speak against a clear cut-off point of 0.5, given that even though $P(f|G)$ was 0, 8 the assertability ratings were not close to ceiling[25]. Given that in this final study we consider different conditional probabilities $P(f|G)$, the results should provide us with a clearer picture of whether the cut-off approach or the gradual change approach defended here come closer to reality.

In this last study, we used the same set-up as in the first two studies. The participants judged the assertability of generic sentences with respect to the two conditions, the non-contrastive condition in which $P(f|G) = P(f|Alt(G))$ and the contrastive condition in which $P(f|Alt(G)) = 0$. The only difference is that now we varied $P(f|G)$ between participants.

As Study 3 was a follow-up to the first two studies, this time we assumed from the start that there will be two groups of participants. Participants that do not take alternatives into account when evaluating the generic sentence (we will refer to this group as *noCon*) are predicted to use the conditional probability of the feature f given the group G as measure of the assertability of the generic sentence. In this case, our theory predicts that in both conditions the assertability of the generic should increase linearly with a growing conditional probability $P(f|G)$. For participants that *do* take the presented alternative into account (group *Con*) the assertability score should depend on $P(f|G)$ and $P(f|Alt(G))$. In the contrastive condition, $P(f|Alt(G))$ is 0 while $P(f|G)$ is not, so again the assertability of the generic sentence should grow linearly with the increase in $P(f|G)$. Furthermore, we predict that the assertability ratings for this

[25]Cohen could argue that this is because some or all of the participants applied the relative reading of generics. However, notice that even after we split participants into groups according to whether they saw a difference between the two conditions, those that did not see a difference still did not give a ceiling assertability score to the generic sentence. Furthermore, in the relative reading, Cohen would predict that still the generic should be completely assertable in the contrastive condition and completely unassertable in the non-contrastive condition, which is again not what we found.

condition should overall be slightly higher (approximately 0.5 points) for the *Con* group than for the *noCon* group[26]. In the non-contrastive condition, both $P(f|G)$ and $P(f|Alt(G))$ are identical so the contingency of the sentence is 0. In this case, for the *Con* group there should be no effect of proportion on the assertability of the generic sentence—the assertability score should be the same independent of $P(f|G)$.

4.5.1. Method

4.5.1.1. Materials

This study had the same design as Study 1, but now we collected data for different proportions with which the animals possessed the relevant color feature. We used four proportions: 54, 68, 80, and 92%[27]. Furthermore, we also varied the distribution of the feature among the 25 animals that were shown to the participants: for each condition we used 3 pictures with different, randomly selected distributions of the feature over the presented animals.

Each participant had to make three judgments: she saw one picture in the contrast condition, one picture with the no contrast condition and one filler, all using the same frequency for the distribution of the feature. Each animal species was shown once. The order of the contrast/no contrast question was randomized, the filler was always shown as the third and last question[28].

4.5.1.2. Participants

Participants were again recruited via Prolific.ac with the same criteria. Four hundred and one participants completed the task. Twenty participants were excluded because they gave inadequate responses to the filler items (score above 1.5). Six further participants were excluded because they gave all three conditions a score 0. Three hundred and seventy-five participants were thus included in the analyses reported below: 97 for frequency 54%, 89 for frequency 68%, 94 for frequency 80%, and 95 for frequency 92%.

4.5.2. Results

As stated above, in this study we distinguish two groups of participants: group *Con* contains participants that found the generic more assertable in the contrastive condition than in the non-contrastive condition; participants in group *noCon* did not give a different score in the two conditions. We split the participants into these two groups using the same criteria as we used in Study 1. There were 135 participants (36%) who gave a higher score in the contrastive condition (*group Con*). When collapsing across different proportions, this group gave a mean score 3.69 (SD 0.97) in the contrastive condition and a mean score 2.0 (SD 1.21) in the non-contrastive condition. There were 209 participants (55%) who gave the same score in the two conditions (*group noCon*). This group gave a mean score 3.2 (SD 1.26) in the contrastive and a mean score 3.18 (SD 1.25) in the non-contrastive condition. Finally, there were 18 participants (9%) who gave a higher score in the non-contrastive condition. The table in **Figure 5** shows the results for the different probabilities split up according to the two groups that we distinguish.

To test our predictions, we conducted a Bayesian ANOVA with condition (contrastive vs. non-contrastive) and proportion (as an ordinal variable) as independent variables for each group separately. To evaluate whether a certain variable has an effect on the given scores, we compared a model including this effect with a model excluding this effect. For the group that gave the same score to both conditions (*group noCon*), we predicted an effect of proportion—the scores should linearly increase with increasing proportions. In the ANOVA analysis, we observed modest evidence against the effect of condition ($BF_{Inclusion} = 0.2$, given by the definition of the group), strong evidence for the effect proportion ($BF_{Inclusion} = 13$), and strong evidence against the interaction of condition and proportion ($BF_{Inclusion} = 0.02$). Thus, we do observe an effect of proportion. However, while the participants did give a higher score with increasing proportions, this increase does not seem to be equally present for all proportion steps. A *post-hoc* test comparing each proportion to the other ones showed that scores given for proportion 54% were not different from scores given for proportion 68% ($BF_{10,U} = 0.16$), and scores given for proportion 80% were not different from scores given for proportion 92% ($BF_{10,U} = 0.22$); for the other proportion pairs, we had evidence for the difference in scores. Thus, participants here did not seem to care about the difference between the lowest two proportions and the highest two proportions, exhibiting rather behavior that would correspond to there being some sort of threshold between $P(f|G) = 68\%$ and $P(f|G) = 80\%$.

For the group that gave the contrastive condition a higher score than the no contrast condition (*group Con*), we predicted an interaction between condition and proportion: the scores given by participants should linearly increase with increasing proportions in the contrastive condition, but they should be the same across proportions in the no contrast condition. In the ANOVA analysis, we observed extreme evidence for the effect of condition ($BF_{Inclusion} = \infty$), inconclusive evidence for presence or absence of the effect of proportion ($BF_{Inclusion} = 0.8$) and modest evidence against the interaction of condition and proportion ($BF_{Inclusion} = 0.2$). Hence, based on our analysis, here the predictions were not borne out—the effects of condition and proportion did not clearly interact. When inspecting averages for each proportion in the two conditions, there *does* indeed seem to be a gradual increase of the scores in the contrast condition in this group, whereas in the no contrast condition there seems to only be a random fluctuation of the scores. But even if we focus only on

[26]The reason for this is a difference in how $P(f|G)$ counts for assertability for participants that take alternatives into account and those that don't. The assertability rating of a participant that doesn't take alternatives into account in the condition where 80% of the animals carries the relevant feature, for instance, should be $P(f|G) * 5 = 4,0$. But a participant that takes alternatives into account should give in the contrastive condition a rating of $\frac{P(f|G)+1}{2} * 5 = 4,5$.

[27]All sample-pictures contained 25 animals of one species (see **Figure 3**). Thus, for example, in the contrastive condition a proportion of 54% means that 14 out of 25 animals in the sample from Genovesa have the property and none of the animals in the sample from Marchena. In the non-contrastive condition in both samples 14 out of 25 animals would have the property.

[28]As a consequence, the trials using 80% were a complete replication of the first study. We will come back to this in the discussion of the results.

		group Con			group noCon		
condition	$P(f\|G)$	Mean	SD	N	Mean	SD	N
contrast, $P(f\|Alt(G)) = 0$	54%	3.34	1.04	37	2.80	1.01	49
	68%	3.67	0.69	35	2.81	1.32	43
	80%	3.79	1.20	32	3.43	1.22	58
	92%	4.06	0.81	31	3.59	1.30	59
no contrast, $P(f\|Alt(G)) = P(f\|G)$	54%	1.71	1.15	37	2.78	1.00	49
	68%	2.20	0.91	35	2.77	1.27	43
	80%	1.89	1.40	32	3.41	1.21	58
	92%	2.27	1.38	31	3.58	1.32	59

FIGURE 5 | Results of study 3.

the judgments for the contrastive condition, there is no evidence for an effect of proportion. It seems like the increase in scores was not consistently present for all participants (see **Figure 6** for a depiction of the individual scores)[29].

Because the proportion with $P(f|G) = 0.8$ is the same frequency of *f*'s given *G*'s that was used in Study 1, we can compare the results for participants that saw this proportion (*N* = 94) with the results obtained in Study 1. For this group, the mean score in contrastive condition was 3.50 (SD 1.25), whereas the mean score in non-contrastive condition was 2.88 (SD 1.47). When split into groups, there were 32 participants (34%) who gave the contrastive condition a higher score (difference more than 0.5) than the non-contrastive condition and 58 participants (61%) who gave them the same score (difference <0.5). Both the averages and the proportions of participants in each group are close to what we observed in Study 1. Hence, these findings are robust.

4.6. General Discussion

All three studies that we reported on support Hypothesis 1: the assertability of a generic sentence *"Gs are f"* depends on the conditional probability of the feature *f* given salient alternatives G' of *G*. We also found evidence for the type of dependency predicted by our proposal made in section 3.2: if the feature *f* is much more frequent given *G* than given the alternative G', then the assertability of the generic improves. Study 1 and study 2 did not support the exact assertability scores predicted by the theory, but as discussed in section 4.4, this might have to do with the particular methodology we used. In particular, our proposal for transformation of the scores in our task to those predicted by the theory might not be accurate.

With study 3, we wanted to investigate whether the predicted dependency on the absolute probability of *f* given *G* is also supported by empirical results. Based on the discussion in section 4.4, we now immediately distinguished two groups within the participants: group *Con* consisted of participants that judged the generic more assertable in the contrastive condition, while in group *noCon* were those participants that gave the same scores in the two conditions.

For the group *noCon*, the results of study 3 supported *a* dependence of assertability on proportion: the assertability increased with the probability $P(f|G)$, independent of condition. But, as discussed above, we could not support the predicted linear increase in assertability that Hypothesis 2 predicts. Instead, there was some evidence for an assertability threshold between the second and third condition of proportion. This provides some evidence for threshold theories like the one of Cohen (1999), though the value of the threshold clearly seems to differ from the 50% threshold that Cohen proposes. Also the values below the threshold are not what one normally would expect. Even in the conditions with $P(f|G) = 54\%$ and $P(f|G) = 68\%$, the generics still where not clearly rejected, but on average still marginally assertable. We need more empirical data, also for different conditions of proportions to be able to say whether we should prefer a threshold account and what form exactly it should take.

For group *Con* the data did not support an interaction between condition and proportion. Note that the mean assertability score given to the generic in the contrastive condition did steadily increase with growing conditional probability of the feature *f* given the group *G*, and in a rate that comes close to what is predicted by the theory. However, statistically the result was not significant. Here, either the theory is wrong or perhaps our experiment was not tapping into the interpretation/significance of alternatives clearly enough to reliably detect the difference. One reason for this could be that this effect (i.e., the increase in scores due to increasing $P(f|G)$) is rather small, so our sample size of approximately 30–35 participants in each group is not large enough to detect it. In this connection, notice also the surprising low assertability ratings of group *Con* for the non-contrastive condition. The theory predicts an assertability value of 0 in this case, independent of $P(f|G)$, which should correspond to a score 2.5 on the scale the participants saw in our study (with our transformation). However, in study 1 and for all four proportions in study 3, the given assertability score was lower than that and varied quite a lot. We already discussed in section 4.4 that a possible explanation might lie in the way people interpreted the scale on which they gave their judgments.

[29]The reader might notice that in the 54% condition only one participant had a very large difference - 5, and there is no other participant in other proportions with such a large difference between the contrastive and the non-contrastive condition. One might think that maybe this participant is the reason why we do not observe an effect of proportion. But excluding this participant does not affect the results.

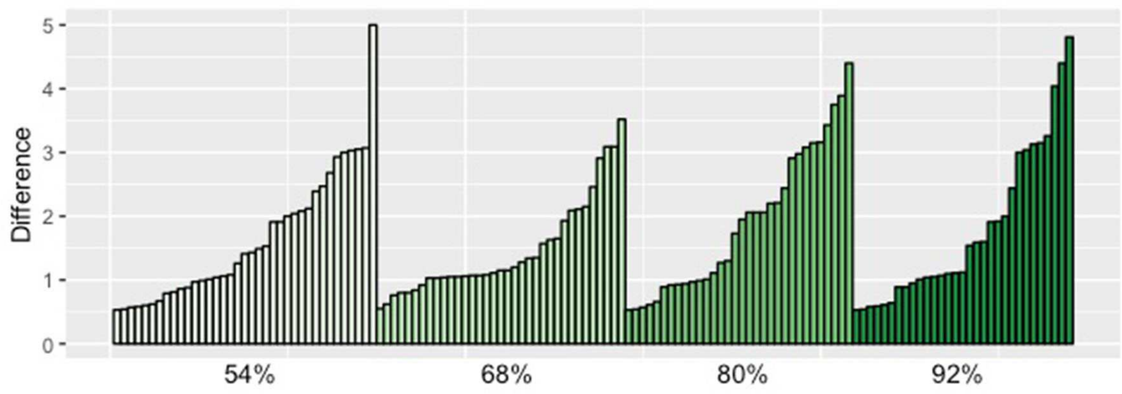


FIGURE 6 | This plot depicts the difference between contrastive and non-contrastive condition (on the Y axis) for each of the 135 participants of the Con group (on the X axis). We grouped the participants by the proportion that they saw. We can see that it is not the case that there are mostly higher scores for higher proportions. NB: each participant saw only one proportion.

Let us turn to the relevance of the data from the group *Con* for the cut-off point hypothesis built into theories like the one proposed in Definition 4 in contrast to the gradual increase in assertability that Hypothesis 2 predicts. As discussed above, for the group *noCon* there was some evidence for a cut-off point between $P(f|G) = 0.68\%$ and $P(f|G) = 0.8\%$. In contrast, for the group *Con* we do not see the same "jump" in assertability ratings between proportions. Instead, as discussed above, at least in terms of just the means there appears to be a linear increase of assertability in the contrastive condition, as predicted by definition 5. From a theoretical point of view this observation is rather difficult to make sense of. Why should there be a cut-off point in case no alternatives are taken into account, while assertability increases linearly in case alternatives do matter? Of course, we could easily propose an ambiguity with two possible readings of generics; one with threshold, one without. But that seems to be an awfully arbitrary difference between two readings of the same sentences. Before we take such a theoretical step we need more evidence that this difference is real. To conclude, our results do not support a clear threshold account, as, for instance, defended in Cohen (1999). But also the linear increase of assertability with growing $P(f|G)$ that Hypothesis 2 predicts is not completely supported by our data.

Finally, there is one more curious feature of the behavior of participants in study 3. Even though the few datapoints we recorded do not allow us to test for it, notice that the size of group *noCon* appears to increase with growing $P(f|G)$. Using the terminology of our proposal, the higher the absolute probability of f given G the less relevant alternatives to G seem to be. There is some evidence from related domains, as studies of causal judgments, showing that actually $P(f|G)$ counts more for the assertability of such judgements (Wasserman et al., 1993; Anderson and Sheu, 1995). Using a measure that takes this into account and, for instance, weighs $P(f|G)$ more the larger this factor is, could explain the tentative observation just made. The higher $P(f|G)$, the less the contrastive value $P(f|Alt(G))$ would count for assertability, and, hence, the smaller the difference between the contrastive and the non-contrastive condition. Consequently, more people would look like belonging to the group *noCon* instead of the group *Con*. Thus, if this tentative observation just made could be supported by a study suitable to test it, it might give us an important hint for how to improve the proposal made here.

Part of the problems we have with supporting the proposal tested here can be probably traced back to shortcomings with the particular experimental setting used here. We already mentioned in section 4.4 the issue with translating the experimental results into the scale of values predicted by the theory. One might be tempted to say that we should not aim at predicting exact assertability values. It is rather unusual for experimental psychology to formulate predictions in terms of specific scores as we did here, because it is assumed that there is too much uncertainty about what people are doing to have such precise predictions; traditionally, only presence or absence of differences between conditions is tested instead. However, we believe that formulating and testing more specific numerical predictions is a good way to reduce the gap between theories like the one about the meaning of generics presented here and experimental findings with human participants. But we also realize that methodologically this presents a number of challenges that we haven't solved completely yet.

Another major issue with the setup we used is that it does not model sequential learning. A central idea of the theory proposed here is that assertability of a generic sentence is equated with the strength of association built based on the frequency with which the agent observed members of a group carrying a particular feature. However, we did not allow the learning of the association to observe these occurrences sequentially. Instead, we just gave the participants of the studies the information in one batch. But probably the limitation of the setup that had the most effect on the results obtained is the lack of control or insight in what the participants of the studies took the relevant alternative set to be. We assumed that the particularly setting used would entice the participants to take the corresponding

alternative sets	$P(f\|Alt(G))$		Assertability	
condition	contrastive	non-contrastive	contrastive	non-contrastive
all animals	≈ 0.0001	≈ 0.0001	≈ 0.8	≈ 0.8
other frogs	≈ 0.15	≈ 0.25	≈ 0.65	≈ 0.55
Marchena frogs	$= 0$	$= 0.8$	$= 0.8$	$= 0$

FIGURE 7 | Assertability values for different alternative sets.

species from Marchena, the species the other sample in each picture was from, as the only alternative to the target group: the species from Genovesa. But nothing in the experimental setting used made sure that this was indeed the case. The participants could have taken all kinds of alternative sets into account. Take, for instance, the example from the questionnaire used given in **Figure 3**. Maybe some participants of the study did indeed take the Tree Frogs from Marchena to be the only relevant alternative. But some might also have compared Tree frogs from Genovesa with what they know about frogs in general. Or they even considered all animals as possible alternatives. What they chose to be the relevant alternatives has, according to the approach tested here, a huge effect on how assertible they considered the given generic sentence about Tree frogs from Genovesa. In fact, this could account to a large extent for the huge variation we observed in the assertability judgements. Let us, as an example, just consider the alternative sets just mentioned and calculate the predicted assertability of the generic *Tree Frogs from Genovesa have yellow dots*. Thus, let $Alt_1(G)$ be the set of all animals, $Alt_2(G)$ be the set consisting only of frogs, and $Alt_3(G)$ be only Marchena Tree Frogs. First, we need to make assumptions about the prior probability of having yellow dots for each of the three potential sets of alternatives—again, this is something that different participants have different opinions about. Let's suppose for the moment that animals in general have only very rarely yellow dots, i.e., $P(f|Alt_1(G)) = 0.0001$, frogs in general, however, tend to have yellow dots much more often, i.e., $P(f|Alt(G_2)) = 0.2$. The participants won't have a prior for $Alt_3(G)$ because this is a novel species for them. Based on these priors, we can now calculate $P(f|Alt(G))$ after the participants saw the picture given in **Figure 3**. This information will hardly change anything for how probable one considers it that animals in general have yellow dots. But it will lower the probability for frogs having yellow dots, lets assume $P(f|Alt_2(G)) = 0.15$. $P(f|Alt_3(G))$ will be 0, based on the information in the picture. After seeing the non-contrastive counterpart of the picture in **Figure 3** the respective probabilities $P(f|Alt(G))$ might be those given in **Figure 7**. These values would result in the assertability values given in the last two columns of the table. As the reader can see, there are huge differences between the various assertability values. For instance, a speaker who takes all animals to be relevant alternatives to the observed species would not see a detectable difference between both conditions, but would take the assertability in both cases to equal the conditional probability $P(f|G)$. Taking a smaller set of alternatives results in some difference between both conditions and a generally lower assertability in both conditions. Taking actually only the species from the alternative island to be a relevant alternative results in the extremely different assertability values that we expected.

This shows first of all that the distinction between participants that do and participants that don't take alternatives into account, which lies at the bottom of the way we analyzed the data of the second study, is not the only way to explain the substantial group of participants that don't see a difference between both conditions. These might also be participants that just consider a very general set of alternatives. Second of all, we have here a way to explain the substantial variation in the data from the perspective of the proposal made. It also points, as said at the beginning of this discussion, to a major weakness of the experimental setup used here. In order to truly test the proposal at hand we need to either control, or probe what the participants of the experiments take to be the relevant alternatives. This will be a focal point of our future work.

5. CONCLUSIONS

The main goal of this paper was to explore and defend a statistical approach to the meaning of generic sentences. Such approaches are in discredit at the moment, because of the various shortcomings of the majority rule, which is the most popular statistical approach to generics on the market at present. However, we think that there is a vast variety of different statistical approaches to the meaning of generics that have a lot of promise. In this paper, we discussed in particular that by taking into account various notions of alternatives for the interpretation of generic sentences, many shortcomings of the majority rule can be overcome. In particular, we argued that alternatives are relevant to the meaning of generics in three different ways. We have seen that alternatives of the property f that the generic ascribes to group G matter, as well as the alternative causal background factors. Finally, we saw that also alternatives to the group G matter for the acceptability of the generic. This has led us to a first and preliminary formal description of the meaning of generic sentences, given in Definition 3.

We then zoomed in on the alternatives to the group G the generic is ascribing some property f to. We motivated the relevance of these alternatives for the meaning of generics by linking this meaning to associative learning. Building on theories of learning from psychology, we formulated a new and final version of our approach. According to this proposal, essential for the assertability of a generic sentence *Gs are f* is how distinctive the feature f is for the group G. We have motivated this approach on the one hand by showing that it can account for many problematic examples in the literature, and on the other hand by showing that such an approach can be motivated by considerations for the psychology of learning and results on the link between statistical information and causal dependence. In short, distinctiveness matters for the assertability of generic sentences, because this condition is essentially linked to how we learn about causal dependencies in the world. This proposal differed from the approach we formulated at the end of the first part of the paper in two important respects. First of all, it predicts the assertability of generics to come in degrees. More concretely, this means that our proposal does not assume strict cut-off points for the truth or assertability of generics. Secondly, the proposal assumes not two, but only one (context-dependent) reading for generic sentences. This reading is the relative reading of Definition 3. The reading can in certain circumstances—if the alternative set the interpreter assumes for G is empty—collaps to the absolute reading of Definition 3.

In the final section of the paper, we reported on three studies that tested our final proposal. In these studies participants were presented with a visual scene and asked to judge the assertability of a generic sentence *Gs are f*. We manipulated the presence of the alternatives and the frequency with which members of group G carried feature f. The results allowed us to confirm the relevance of G-alternatives for the meaning of generic sentences in the population in general. We also observed some evidence for the correlation between assertability of generic sentences and $P(f|G)$. However, not all particular predictions made by the proposal in section 3.2 were borne out.

We also saw that the experiment setting explored here still has a number of shortcomings. Two should be the focus of future work along the lines explored here. First of all, we need to develop an experimental paradigm that allow us to test the link made here between the assertability of generic sentences and learning more directly. In particular, we need to model learning more naturally in the experimental setting. The second is to find a way to gain more insight or control on what the speaker of a generic sentence takes to be the relevant alternatives. As we have seen in the last part of the previous section, assuming that there was a lot of variety of what the participants of the studies took to be the relevant G-alternatives can account for huge variation in the assertability judgements observed. In future work we need to invest in experimental methods that allow us to probe or manipulate these alternatives sets.

Though the most pressing challenges for future work on the topic explored here are arguably methodological in nature—we need a solid empirical basis in order to direct further theoretical work—there are also a couple of interesting theoretical questions that we want to explore in future work. Just to mention one example, we picked contingency to measure associative learning. However, causal impact was not tested and there are also other measures of strength of association discussed in the literature. We should test those as well on the data-set gathered here and compare the predictions made with those of contingency.

DATA AVAILABILITY STATEMENT

The raw data and analysis script are available on https://osf.io/hyt8d/, doi: 10.17605/OSF.IO/HYT8D.

ETHICS STATEMENT

Ethical review and approval was not required for the study on 2022 human participants in accordance with the local legislation and 2023 institutional requirements. The participants provided their informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work for this paper was financially supported by RV NWO Open Competition grant 406.18.TW.007, from Learning to Meaning: a new approach to Generic Sentences and Implicit Biases.

REFERENCES

Anderson, J. R., and Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Mem. Cogn.* 23, 510–524. doi: 10.3758/BF03197251

Asher, N., and Morreau, M. (1995). "What some generic sentences mean," in *The Generic Book*, eds G. Carlson and F. J. Pelletier (Chicago, IL: Chicago University Press), 300–339.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *Q. J. Econ.* 131, 1753–1794. doi: 10.1093/qje/qjw029

Carlson, G. (1977). Reference to kinds in english (Ph.D. dissertation). University of Massachusetts, Amherst, MA, United States.

Cheng, P. (1997). From covariation to causation: a causal power theory. *Psychol. Rev.* **104**, 367–405. doi: 10.1037/0033-295X.104.2.367

Cimpian, A., Gelman, S., and Brandone, A. (2010a). Theory-based considerations influence the interpretation of generic sentences. *Lang. Cogn. Process.* 25, 261–276. doi: 10.1080/01690960903025227

Cimpian, A., Gelman, S., and Brandone, A. (2010b). Generic statements require little evidence for acceptance but have powerful implications. *Cogn. Sci.* 34, 1452–1482. doi: 10.1111/j.1551-6709.2010.01126.x

Cohen, A. (1999). *Think Generic! The Meaning and Use of Generic Sentences.* Stanford, CA: CSLI Publications.

Cohen, A. (2004). Generics and mental representations. *Linguist. Philos.* 27, 529–556. doi: 10.1023/B:LING.0000033851.25870.3e

Cox, A. (2004). *U.S. Poised for Epidemic West Nile year.* Available online at: http://edition.cnn.com/2004/HEALTH/05/03/wnv.outlook/index.html (accessed January 28, 2020).

Dahl, Ö. (1975). "On generics," in *Formal Semantics of Natural Language: Papers From a Colloquium Sponsored by the King's College Research Centre, Cambridge*, ed E. L. Keenan (Cambridge: Cambridge University Press), 99–111.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *J. Math. Psychol.* 47, 109–121. doi: 10.1016/S0022-2496(02)00016-0

Declerck, R. (1986). The manifold interpretations of generic sentences. *Lingua* 68, 149–188. doi: 10.1016/0024-3841(86)90002-1

Gelman, S. (2004). "Learning words for kinds: generic noun phrases in acquisition," in *Weaving a Lexicon*, eds D. G. Hall and S. R. Waxman (Cambridge, MA: MIT Press), 445–484.

Gelman, S., and Bloom, P. (2007). Developmental changes in the understanding of generics. *Cognition* 60, 115–140. doi: 10.1016/j.cognition.2006.09.009

Gilson C., and Abelson, R. P. (1965). The subjective use of inductive evidence. *J. Pers. Soc. Psychol.* 2, 301–310. doi: 10.1037/h0022268

Gluck, M. A., and Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *J. Exp. Psychol.* 117, 227–247. doi: 10.1037/0096-3445.117.3.227

Grimm, S. (2009). "Number marking and individuation: a view from dagaare," *Proceedings of Sinn und Bedeutung 14*, eds M. Prinzhorn, V. Schmitt, and S. Zobel (Vienna), 168–184.

JASP Team (2018). *JASP (Version 0.9) [Computer Software]*. Available online at: https://jasp-stats.org/

Lawler, J. M. (1973). Studies in english generics. *Univ. Michigan Pap. Linguist.* 1, 1–184.

Leslie, S. J. (2008). Generics: cognition and acquisition. *Philos. Rev.* 117, 1–47. doi: 10.1215/00318108-2007-023

Murphy, G. L. (2004). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., and Medin, D. L. (1985). The role of theories in conceptual conherence. *Psychol. Rev.* 92, 289–316. doi: 10.1037/0033-295X.92.3.289

Newel, B., Lagnado, D., and Shanks, D. (2007). *Straight Choices. The Psychology of Decision Making*. Hove; New York, NY: Psychology Press.

Nickel, B. (2012). "Dutchmen are good sailors: generics and gradability," in *Genericity*, eds Mari et al. (Oxford: Oxford University Press), 390–405.

Palan, S., and Schitter, C. (2018). Prolific. ac–A subject pool for online experiments. *J. Behav. Exp. Finance* 17, 22–27. doi: 10.1016/j.jbef.2017.12.004

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. New York, NY: Cambridge University Press.

Prasada, S. (2000). Acquiring generic knowledge. *Trends Cogn. Sci.* 4, 66–72. doi: 10.1016/S1364-6613(99)01429-1

Prasada, S., and Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition* 99, 73–112. doi: 10.1016/j.cognition.2005.01.003

Rescorla, R., and Wagner, A. (1972). "A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement," in *Classical Conditioning II*, eds A. Black and W. Prokasy (New York, NY: Appleton-Century-Crofts), 64–69.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *J. Compar. Physiol. Psychol.* 66:15. doi: 10.1037/h0025984

Skyrms, D. B. (1980). *Causal Necessity. A Pragmatic Investigation of the Necessity of Laws*. New Haven, CT; London: Yale University Press.

Stolz, T. (2001). Singulative-collective: natural morphology and stable classes in welsh number inflexion on nouns. *Sprachtypol. Univ.* 54, 52–76. doi: 10.1524/stuf.2001.54.1.52

Tessler, M., and Goodman, N. (2019). The language of generalization. *Psychol. Rev.* 126, 395–436. doi: 10.1037/rev0000142

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

van Rooij, R., and Schulz, K. (2019). A causal power theory of generics. *Topoi*. doi: 10.1007/s11245-019-09663-4

van Rooij, R., and Schulz, K. (2020a). Generics and typicality: a bounded rationality approach. *Linguist. Philos.* 43, 83–117. doi: 10.1007/s10988-019-09265-8

van Rooij, R., and Schulz, K. (2020b). A causal semantics for IS generics. *J. Semant.* 37, 269–295.

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., and Baker, A. G. (1993). Rating causal relations: the role of probability in judgments of response-outcome contingency. *J. Exp. Psychol.* 19, 174–188. doi: 10.1037/0278-7393.19.1.174

Yuille, A. (2006). "Augmented Rescorla-Wagner and maximum likelihood estimation," in *Advances in Neural Information Processing Systems 18*, eds Y. Weiss, B. Schoelkopf and J. Platt (Cambridge, MA: MIT Press), 1561–1568.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 14 January 2021
doi: 10.3389/fcomm.2020.461553

Scalar Implicature, Hurford's Constraint, Contrastiveness and How They All Come Together

*Satoshi Tomioka**

Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, United States

Disjunction with two scalar items, such as *some or all of the books*, has been regarded as evidence for the grammatical theory of scalar implicatures (e.g., Chierchia et al., 2012). Hurford's Constraint (Hurford, 1974) provides that disjuncts are banned from having an entailing relation, and to make such a disjunction comply with Hurford's Constraint, the meaning of *some* must be locally strengthened. Interestingly, however, the order of disjoined scalar items is not free, as noted by Singh (2008). The order in which a weaker scalar item comes first followed by its stronger scalar mate is better than the other order. I present an analysis of this ordering restriction based on the novel observation that the restriction is not only found in disjunction but in contrastive environments in general. I propose that contrasting a linguistic expression requires a "contrast antecedent," which must elicit a set of mutually exclusive alternatives that includes the meaning of the contrasted expression. It will be demonstrated how the mutual exclusivity requirement presents a principled explanation for the ordering asymmetry as well as Hurford's Constraint itself, which indicates that the root of the constraint is not in disjunction but in contrastiveness. One of the indispensable ingredients in the proposal is the grammatical/conventional generation of scalar implicatures, as the strengthened meaning must be the basis of alternatives. The paper also provides a speculative analysis of *only*, in which I suggest that the process of exhaustification in the grammatical theory of scalar implicatures should not be characterized as the implicit *only*, the semantic contributions of which are more different than commonly assumed.

Keywords: scalar implicature, disjunction, mutually exclusive, exhaustivity, contrastive

Edited by:
Sophie Repp,
University of Cologne, Germany

Reviewed by:
Benjamin Spector,
École Normale Supérieure, France
Stefan Hinterwimmer,
University of Wuppertal, Germany

****Correspondence:***
Satoshi Tomioka
stomioka@udel.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *22 March 2020*
Accepted: *09 December 2020*
Published: *14 January 2021*

Citation:
Tomioka S (2021) Scalar Implicature, Hurford's Constraint, Contrastiveness and How They All Come Together. Front. Commun. 5:461553. doi: 10.3389/fcomm.2020.461553

1. HURFORD'S CONSTRAINT: WHAT IT IS AND WHY WE CARE

Hurford (1974) noted that there is a felicity constraint that bans entailing disjuncts, as exemplified in (1).

(1) a. Anna lives in Seattle or in California.
b. #Anna lives in Los Angeles or in California.

Living in Los Angeles entails living in California, so they cannot be placed side by side in disjunction. No such ill effects are found in (1a), in which no entailment relation holds[1].

Hurford further notes that *A or B* and *A and B* can be disjoined.

(2) Inmates may smoke or_1 drink, or_2 both. = (20) in Hurford (1974, p. 410).

If both instances of *or* were to be interpreted as logical disjunction, this sentence is incorrectly predicted to be infelicitous because *inmates may both smoke and drink* entails *inmates may smoke or drink*. Hurford's conclusion is that the first instance (= or_1) is exclusive. In other words, it means, "Inmates may smoke or drink but not both."

As noticed by Gazdar (1979), however, this alleged strengthening is not limited to disjunction. A variety of "scalar mates" of the *Horn Scales* of Horn (1972) can be disjoined, as shown below.

(3) a. Erica visited France or Italy, or (both) France and Italy.
b. Anna ate some or all of the cookies.
c. Is Maria's academic record good or outstanding?
d. Is wearing a helmet recommended or required?

Logically, eating all of the cookies entails eating some of the cookies. The relations in (3cd) may be a little less straightforward, but one can still argue that entailment is involved there as well. Nonetheless, these sentences do not show the expected effects of Hurford's constraint violations. To make the disjunctions in (3) comply with Hurford's Constraint, it is necessary to strengthen the meaning of the scalar expressions that are in the entailed propositions, as indicated in (4).

(4) a. Erica visited France or Italy (but not both), or France and Italy.
b. Anna ate (only) some or all of the cookies.
c. Is Maria's academic record (just) good or outstanding?
d. Is wearing a helmet (only) recommended or required?

Crucially, the silent addition of those underlined expressions in the parentheses must take place before the meaning of the disjunctive structure is computed. This is the point that is at odds with the traditional Gricean approach to scalar implicatures. As a kind of conversational implicature, scalar implicature is standardly assumed to be generated after the calculation of the semantic meaning of a whole sentence is completed.

Hurford's Constraint with disjoined scalar items, as well as other instances of embedded implicatures, has led to the emergence of the grammaticist or conventionalist approach to scalar implicature, represented by Chierchia (2006), Fox (2007), and Chierchia et al. (2012)[2].

(5) a. A scalar item generates scalar alternatives based on its meaning (using the Horn Scale).
b. The exhaustivity operator, O_{ALT}, can be inserted at any sentential level.
c. This operator negates all of the non-weaker alternatives to the denotation of its complement, and the negated proposition is added to the denotation of its complement[3]. In other words, O_{ALT}(S) = S and all the propositions not entailed by S are false.
d. The addition of the negated proposition takes place at the level in which O_{ALT} appears.

It appears that what O_{ALT} does is not too far from the semantic effect of the adverb *only*, and O_{ALT} is often described as the silent version of it. Chierchia et al. (2012) suggest that the silent version of *only* is independently needed to derive exhaustive answers to questions. When someone utters, "Anna introduced [ERIC and FRED]$_F$ to Maria" as an answer to the question, "Who did Anna introduce to Maria?", it is naturally interpreted that Anna introduced Eric and Fred and no one else to Maria. According to Chierchia et al. (2012, p. 8), "What may be going on is that focus activates alternatives; active alternatives must be put to use and one option is via a covert occurrence of *only*. We may, then, assume that something very similar happens in the case of scalar alternatives."

The grammar of O_{ALT} presents a very straightforward representation of each sentence in (3) which does not violate Hurford's Constraint. Here is how (3b) complies with the constraint.

(6) a. [[O_{ALT} [Anna ate some (of the cookies)] or [(she ate) all of the cookies]].
b. [O_{ALT} (Anna ate some of the cookies) $\rightsquigarrow$ Anna ate some of the cookies, and Anna did not eat all of the cookies.
c. [[O_{ALT} [Anna ate some (of the cookies)] or [(she ate) all of the cookies]] $\approx$ Anna ate only some of the cookies or all of the cookies.

2. ORDERING ASYMMETRY

While two scale-mate expressions can be disjoined, the order of the expressions is not free. Singh (2008) notes that a semantically stronger item cannot precede its weaker counterpart. I will place

[1] Singh (2008) points out that we need a stronger constraint.

(i) a. Bertha drives a pick-up truck or an SUV.
b. #Bertha drives a pick-up truck or a Ford.

There is no entailment relation between *Bertha drives a pick-up truck* and *Bertha drives a Ford*, which are merely consistent with each other. Singh strengthens Hurford's Constraint as: # *X or Y* if *X* and *Y* are mutually consistent (Singh, 2008, p.252). For the purpose of the discussion in this current paper, however, we will not make reference to this stronger version of Hurford's Constraint.

[2] Of the grammaticist approach, I focus primarily on the proposal that makes use of a sentential exhaustivity operator, rather than other grammaticist analyses without such an operator (e.g., Levinson, 2000; Chierchia, 2004).

[3] Strictly speaking, the operator negates all the non-weaker alternatives that are "innocently excludable" (see Fox, 2007 for discussion). For the purpose of this paper, the simpler version is sufficient.

in all of the examples below, but our judgments may fluctuate (and Fox and Spector, 2018 report that their corpus search produces some sizable number of exceptions, which they try to explain).

(7) a. #Erica visited [(both) France and Italy] or [France or Italy].
b. #Anna ate all of the cookies or some of the cookies.
c. # Is Maria's academic record outstanding or good?
d. # Is wearing a helmet required or recommended?

These examples become acceptable if *only* or *just* is added to the second disjuncts.

(8) a. Erica visited [(both) France and Italy] or just [France OR Italy].
b. Anna ate all of the cookies or only some of the cookies.
c. Is Maria's academic record outstanding or just good?
d. Is wearing a helmet required or only recommended?

The two analyses of this asymmetry, namely Singh (2008) and Fox and Spector (2018), both have the following characteristics.

(9) a. The contrast has its root in the way the silent exhaustive operator (O_{ALT}) is licensed.
b. The licensing of O_{ALT} is sensitive to structure (= disjunction) and its impact on the semantic environment.
c. In other words, [(O_{ALT}) A or B] is legitimate, but [A or (O_{ALT}) B] is not.

Let us quickly review the gist of Fox and Spector's (2018) analysis. Compare the two disjunctive structures (10ab).

(10) a. [O_{ALT} [$_S$ Anna ate some of the cookies]] or [$_S$ Anna ate [all] of them]
b. [$_S$ Anna ate [all] of the cookies] or [O_{ALT} [$_S$ Anna ate [some] of them]]

It should be noted that the presence of O_{ALT} has no effect on the context change potential of the whole sentences in (10). Whether O_{ALT} is attached or not, the assertion of those sentences leads to the elimination from the context set of all the possible worlds in which Anna did not eat any of the cookies. In this sense, the attachment of O_{ALT} is *globally vacuous* in both cases. At the point in which the attachment of O_{ALT} takes place in (10a), however, it is still unknown whether the presence of O_{ALT} in the first clause is vacuous or not. The first clause could have ended as an independent sentence, and in such a case, O_{ALT} is not vacuous. On the other hand, the presence of O_{ALT} in (10b) is doomed to be vacuous when the attachment takes place. When the first sentence is "Anna ate all the cookies" and it is followed by the disjunction "or," we need not wait till the meaning of the entire sentence is computed in order to know that O_{ALT} is vacuous. Thus, the presence of O_{ALT} is *incrementally vacuous* in (10b) but not in (10a). The incrementally vacuous use of O_{ALT} is not allowed, and as a consequence, the second disjunct of (10b) cannot furnish O_{ALT}, which leads to a violation of Hurford's Constraint.

What I am about to propose for the ordering asymmetry is rather different from the previous analyses. The reason is clear and simple: I begin with an entirely different descriptive generalization of the ordering asymmetry. More concretely, I will show that the asymmetry is not limited to disjunctive structures, and that the relevant notion is contrastiveness, which includes disjunction but other constructions as well.

3. ORDERING ASYMMETRY IN CONTRASTIVE ENVIRONMENTS

Hurford's Constraint has been regarded as a constraint on disjunction, but it is not hard to find a similar pattern in other cases as well. Giorgio Magri (personal communication) pointed out, for instance, that the contrastive conjunction with *but* seems to impose the same restriction[4].

(11) a. #Adam was born in Paris but Bill in France.
b. #Adam was born in France but Bill in Paris.

(12) a. #Adam has a dog but Bill has a German Shepard.
b. #Adam has a German Shepard but Bill has a dog.

Perhaps not surprisingly, contrasting scalar items works in the same way as well: two scalar items can be contrasted with each other even when one logically entails the other, provided that (i) the entailed item has its meaning strengthened, and (ii) the logically entailed item linearly precedes the entailing item[5].

(13) a. Adam did some of the homework, but Bill did all of it.
b. Adam loves [Ann or Sue], but Bill loves [Ann and Sue].
c. Adam's academic record is good, but Bill's is exemplary.

(14) a. #Adam did all of the homework but Bill did some of it.
b. #Adam loves [Ann and Sue] but Bill loves [Ann or Sue].
c. #Adam's academic record is exemplary, but Bill's is good.

It turns out that we do not have to look hard to find other contrastive environments where the same ordering asymmetry is observed. In all of the examples below, the order in which the logically entailing item precedes the entailed item is judged odder than the other order, and the degraded order is saved by the presence of an overt exhaustifying expression, such as *only* or *just*.

[4] I am grateful to Giorgio Magri for sharing data from his unpublished manuscript.
[5] Winterstein (2012) notes the same contrast [2012, (18) and (19)] but not in connection with Singh's (2008) paradigm of disjunction ordering.

(15) a. Anna is more likely to have solved some of the problems than to have solved all of them.
b. # Anna is more likely to have solved all of the problems than to have solved some of them.
c. Anna is more likely to have solved all of the problems than to have solved only some of them.

(16) a. Anna is more likely to become a good scholar than a brilliant one.
b. # Anna is more likely to become a brilliant scholar than a good one.
c. Anna is more likely to become a brilliant scholar than just a good one.

The comparison between (17) and (18) reveals that the critical factor is the linear order, rather than the matrix-subordinate distinction.

(17) a. While/Although Anna read some of the books, Maria read all of them
b. # While/Although Anna read all of the books, Maria read some of them
c. While/Although Anna read all of the books, Maria read only some of them

(18) a. Anna read some of the books while/although Maria read all of them.
b. # Anna read all of the books while/although Maria read some of them.
c. Anna read all of the books while/although Maria read only some of them.

It is unclear whether all of these contrastive environments above have some structural commonality that can form the basis for an explanation of the asymmetry, but the pursuit in that direction is bound to be unsuccessful, as the same ordering asymmetry is found even in dialogs, as illustrated below[6].

(19) A: Some of Professor Smith's students are smart.
B: I disagree! All of them are smart.

(20) A: All of Professor Smith's students are smart.
B: I disagree! #Some of them are smart.
B': I disagree! Only some of them are smart.

(21) A: The food at that new restaurant was decent.
B: I disagree! I went there last week, and it was excellent.

(22) A: The food at that new restaurant was excellent.
B: I disagree! I went there last week, and # it was decent.
B': I disagree! I went there last week, and it was only decent.

These dialog examples are particularly problematic for the previous analyses of the asymmetry. The infelicitous responses in (20B) and (22B) are independent sentences. When they are uttered in isolation, they are most likely to generate the relevant scalar implicatures as in (23).

(23) a. Some of Professor Smith's students are smart. $\rightsquigarrow$ Some but not all of Professor Smith's students are smart.
b. It (the food at the new restaurant) was decent. $\rightsquigarrow$ It (the food at the new restaurant) was decent but not great.

If these readings are due to O_{ALT}, it is not clear why its presence is blocked in the contrastive environments in (20) and (22).

Let us summarize what we have observed so far. The ordering asymmetry of Singh (2008) is found in contrastive environments beyond disjunction. When two scale-mates are contrasted, the better order is the one in which the semantically stronger one follows the weaker one. The second step is to identify the source of the asymmetry. The data examined above, especially the dialog cases, suggest that it is not about how to regulate the attachment of O_{ALT}. The alternative analysis I explore is based on the following intuition.

(24) When two scale-mates are contrasted, the preceding one must be the right kind of "contrast antecedent" for the following one.

The idea is that the preceding scalar item must "set the stage" for a subsequent contrast. In the maligned *all* - *some* order, for instance, *all* fails to make a necessary preparation for the contrastive use of *some*. In what follows, I will spell out the condition for a good contrast antecedent.

4. CONTRASTIVE FOCUS AND CONTRAST ANTECEDENT

4.1. Strengthening the Condition for Contrastive Focus

The examples that we have examined so far highlight two crucial concepts: focus and contrast. The involvement of focus is obvious

[6]An anonymous reviewer questions the judgment on (20B) and (22B). It is important to note that the infelicity of (20B) is based on the interpretation that B is reporting his belief that some but not all of Professor Smith's students are smart. However, the reviewer is right in that there is one way to make the mention of *some* acceptable in this context. The following is another hypothetical reply to A's statement.

(i) B": Wait, I know that some of them are, but are you sure that all of them are smart?

This reply with rise-fall-rise (RFR) prosody on *some of them are* leads to the "at least some" interpretation, rather than the "some but not all" interpretation. It is a feature of what is often described as "contrastive topic" (Büring, 1997, 2003; Lee, 2007; Tomioka, 2010), which generates a sense of unsettledness, uncertainty, or lack of commitment on the part of the speaker. In this particular example, the speaker indicates by using the contrastive topic prosody that she is not certain about the stronger alternative to "some of them are smart" (i.e., "all of them are"). The end result of this speaker's uncertainty is the "at least some" interpretation. Example (i) shows that, despite its name, contrastive topic is not regulated by our contrastiveness-based condition (the Contrast Antecedent Condition, Section 4.1.). One popular characterization of contrastive topic is that it embodies a particular strategy of answering a QUD (or a sub-question of a QUD). If so, its relation to the contrast antecedent (e.g., the corresponding scalar expression in the previous discourse) is much more indirect. The kind of contrastive dialog that (19) – (22) exemplify involves a more direct contrastive relation, such as denial or correction.

because all the relevant scalar items in the examples above receive focal accents. Focus encompasses a vast range of empirical phenomena, and it does not seem to have a direct connection to Hurford's constraint and its ordering asymmetry.

(25) a. Anna is from $[\text{PARIS}]_F$. So, (it means) she is from $[\text{FRANCE}]_F$.
b. Currently, Anna is in $[\text{FRANCE}]_F$. More concretely, she is in $[\text{PARIS}]_F$.
c. Anna's dream is to live in $[\text{PARIS}]_F$, but she might still be quite content if she lived in $[\text{FRANCE}]_F$.

Being in Paris entails being in France, but such a relation does not seem to negatively affect the focalization pattern in (25), and the ordering does not matter, either. Rooth's (1992) theory of focus intends to capture the generality of focus. As briefly mentioned in connection with the notion of exhaustivity in Chierchia et al. (2012), focusing a linguistic expression elicits a set of alternative denotations of that expression. In Rooth's theory, this set denotation is used in a kind of anaphora resolution scheme: When a linguistic expression α is focused, there must be either (i) another linguistic expression β whose denotation is a subset of the set of alternatives to the meaning of α or (ii) one whose meaning is an element of the set of alternatives to the meaning of α. In (25a), for instance, "Anna is from $[\text{France}]_F$" evokes a set of propositions of the form "Anna is from x." The previous sentence, "Anna is from Paris" can satisfy the condition since it can be an element of that set. It is clear that the ban on an entailment relation is not a part of Rooth's licensing condition, and that it should not be.

What focus provides is a set of alternatives, but this set is not constrained enough to derive the ban on entailment relations and the ordering asymmetry. This is where the second concept, contrast, comes into play. It is worth noting that the sentences in (25) do not have a strong sense of contrast comparable to the examples that we examined in the preceding section. The two sentences in (25a) could jointly answer the Question-under-Discussion, "What is Anna's nationality?", and (25c) can be uttered when someone wishes to know where Anna would like to live.

I admit that it is not always straightforward to give a precise characterization of what counts as "contrastive." There are definitely some expressions or constructions that generate a sense of contrast; disjunction, expressions such as *but, on the other hand, instead, while, although, etc.*, and constructions like comparatives. In case of inter-speaker utterances, we may appeal to the rhetorical relation of contrast in the sense of Kehler (2002) and Asher and Lascarides (2003). In this paper, I primarily use example sentences with overt linguistic cues of contrast, such as disjunction or *but*, except for the few occasions when reference is made to dialog cases. In the dialog examples, the act of "correction" or "challenge" is used to elicit the necessary sense of contrast.

Intuitively speaking, it is fairly natural to suppose that, when two or more expressions are contrasted, their meanings are distinct from each other. For instance, *pick-up truck* and *passenger car* can be contrasted, but *pick-up truck* and *vehicle* cannot. It is also odd to contrast *passenger car* with *BMW* where two properties overlap. The situation is reminiscent of Singh's (2008) stronger version of Hurford's Constraint based on mutual inconsistency (see footnote 1). I propose that contrastive focus demands a more specialized set of alternatives, namely a set of **mutually exclusive** alternatives, and that this stronger notion of alternatives is imposed on the scalar item that comes first[7].

(26) **Contrast Antecedent Condition**
When α is contrastively focused, there must be β that precedes α and generates ALT_β, a set of alternatives for β, such that
(i) it is a subset of the focus semantic value of β,
(ii) its members are mutually exclusive, and
(iii) it includes both the ordinary value of α and that of β.

I take the notion of mutual exclusiveness to be "mutual inconsistency" for propositions and "no mereological overlap" for entities. For other types, their mutual exclusivity is recursively defined: For any expressions α, β of type <a,b>, $[\![\alpha]\!]$ and $[\![\beta]\!]$ are mutually exclusive iff for all x $\in D_a$, $[\![\alpha]\!](x)$ and $[\![\beta]\!](x)$ are mutually exclusive. The term "mutual" should be understood to be strongly reciprocal. For a set to be mutually exclusive, any given pair of its members are mutually exclusive.

The idea of "contrast antecedent" in (26) embodies an unevenness that ultimately leads to the ordering asymmetry. However, the asymmetric aspect of contrast antecedent may be counter-intuitive because we often take it for granted that "contrast" is a symmetric relation. We often make a reciprocal statement like "A and B contrast with each other," and "A contrasts with B" seems equivalent to "B contrasts with A." The symmetry of contrast is also reflected in some linguistic analyses. In Rooth (1992), for instance, contrastively focused items can have a symmetric, inter-dependent relation with each other.

(27) An $[\text{American}]_F$ farmer was talking to a $[\text{Canadian}]_F$ farmer… Rooth [1992, p. 80 (11)]

The focus anaphor introduced with *[American]*$_F$ takes *[Canadian]*$_F$ as its antecedent, and vice versa. However, this symmetry of contrast is based on the notion of contrast as a state of affairs. What about the act of contrasting? First of all, language is a linear system, and therefore it cannot express a symmetry in its purest form. When two expressions are contrasted, one must linearly precede the other in the actual utterance since two expressions cannot be pronounced simultaneously. For this reason, we should not be too shocked to find linear order effects in seemingly symmetric linguistic environments. Take the contrastive use of *but*, for instance. The relevance of contrastiveness in the *but* conjunction is evident, and some have argued (e.g., Sæbø, 2003; Umbach, 2005) that contrast is the essential property that regulates the *but* conjunction. It has been noted (Blakemore and Carston, 2005; Winterstein, 2012 among many others), however, that

[7]Mutually exclusive sets of alternatives have been proposed for the universal free choice item, *cualquiera*, in Spanish by Menéndez-Benito (2010) and for focused adjectival modifiers by Wagner (2012).

there are numerous instances of asymmetric *but* conjunctions where *p but q* and *q but p* are not equivalent. While Kehler (2002) argues that *but* marks either a discourse relation that is symmetric (a CONTRAST relation) or one that is asymmetric (a denial of expectation relation), Winterstein (2012) refutes this ambiguity-based characterization, showing that the distinction is not always straightforward. Particularly relevant is a case in which a contrast is undeniably present but the conjuncts are not symmetric.

(28) *Where will you take your parents next year?*

a. Well, FATHER would like somewhere NEARBY, but MOTHER really wants to go to PARIS.

b. Well, MOTHER really wants to go to PARIS, but FATHER would like somewhere NEARBY.

In both replies, *father* and *nearby* are contrasted with *mother* and *Paris*, respectively. The presence of contrast does not make the two *but* conjunctions equivalent, however. For instance, (28a) can be followed by a statement such as "So, Paris is a possibility." Such a continuation is rather unnatural in (28b). The precise explanation for this contrast is not too important for our current discussion. The lesson we learn from it is that the contrastive *but* can be asymmetric. In addition, a contrastive statement in a dialog presents an even clearer case of asymmetric contrast, as we have seen in the examples in (19) – (22). One can utter a contrastive sentence in connection to a preceding utterance by someone else, and in such a case, contrast is clearly asymmetric. A very short expression of contrast, such as a phrasal disjunction structure uttered by a single speaker, may feel very symmetric, but it is a special case, rather than a general case. In the scheme of (26), *A or B* can be regarded as symmetric (in the sense that it is identical to *B or A*) when A is a good contrast antecedent of B and B would be a good contrast antecedent of A if the order were reversed. Such a case "over-qualifies" with respect to the contrast antecedent condition, and it is special in that sense. I suggest that even in innocuously symmetric-looking contrastive environments, the order of presentation could matter, and that the condition on contrast antecedents captures the impact, one manifestation of which is the ordering asymmetry of scalar items.

4.2. The Contrast Antecedent Condition at Work

We are now ready to examine how the proposed conditions on contrast antecedent can account for the ordering asymmetry. Let us begin with the felicitous *some* - *all* order.

(29) [Anna ate $[\text{some}]_F$ of the cookies] or [Anna ate $[\text{all}]_F$ of the cookies]

$[\![[\text{some}]_F]\!]_f$ is a set of quantificational determiner meanings. While a focus value itself includes any expression of the same semantic type, the actual set of alternatives that are used for comparison (i.e., ALT_{some}) is much more constrained. Relatedly, the issue of scalar alternatives is a hotly debated topic (Horn 1989, Matsumoto 1995, Katzir 2007 among many others). The computation of a scalar implicature is assumed to use a set of scalar alternatives, and making a wrong choice in selecting alternatives can lead to an unattested implicature. There is also a question of how much the scalar alternatives for generating implicatures can influence the choice of focus alternatives. Fox and Katzir (2011) argue that the two sets of alternatives are one and the same, endorsing the structural-complexity-based account of Katzir (2007)[8]. In analyzing scalar contrasts, I partially adopt Fox and Katzir's analysis in that the structural complexity can serve as a restriction on focus alternatives although its role is more limited than what Fox and Katzir envision. For the purpose of dealing with the ordering asymmetry, a relatively informal version of Katzir's constraint is sufficient: For a linguistic expression α, focus alternatives must be structurally at most as complex as α[9].

Turning back to ALT_{some}, it includes only those quantificational determiners the structures of which are at most as complex as *some*, such as *every, no* and probably non-logical lexical quantifiers like *most, many*, etc. Crucially, structurally complex quantifiers (e.g., *almost all, many but not all*) are not included. However, not all of the simplex quantifiers survive if ALT_{some} is to be mutually exclusive. The mutually exclusive set of alternatives contains just two members, {some, no}. Since this set does not include the meaning of *all*, the contrast antecedent condition is still not met.

It is possible, however, to create a larger mutually exclusive set by strengthening the meaning of *some* to *some but not all*. In the scheme proposed by Chierchia et al. (2012), this can be achieved by adjoining O_{ALT} to the first disjunct. In this paper, I will combine O_{ALT} with the scalar item itself for convenience, and this choice should not be regarded as my endorsement of the lexicalist approach to implicatures (e.g., Levinson, 2000). We will revisit the sentential operator analysis of O_{ALT} in section 4.4. With the meaning of *some* strengthened, $ALT_{O[some]}$ becomes {some but not all, no, all}, where *all* is now added since it becomes mutually exclusive with the other two quantifiers[10,11]. Since $ALT_{O[some]}$ includes the ordinary value of the second scalar item, the condition is now met. We should also be reminded that the strengthened meaning of *some* is often taken to be the default meaning (to the extent that some researchers, most notably Levinson (2000), argue that it is lexically encoded), which in turn means that the inclusion of the meaning of *all* in the mutually exclusive set is done fairly easily.

[8]However, Fox and Katzir's (2011) discussion primarily concentrates on association-with-focus cases.

[9]Slightly more formally: for any pair of two structures, S, S', S' is at most as complex as S in a context C if S' can be derived from S by successive replacements of subconstituents of S with elements of the substitution source for S in C [Fox and Katzir, 2011, (34)].

[10]I assume the presuppositional reading of *all* where its argument cannot be the empty set. This assumption is conventionally adopted in the studies of scalar implicatures.

[11]Since the structural complexity of *some* increased to [O_{ALT} *some*], O_{ALT} can in principle contain other quantifiers of the form [O_{ALT} *Q*] where *Q* is a simplex quantificational determiner. However, the other candidates, such as [O_{ALT} *most*], are not mutually exclusive with [O_{ALT} *some*]. Moreover, there is an additional constraint on the set based on relevance of alternatives, as we will discuss later in this section.

On the other hand, the *all* – *some* order cannot satisfy the condition, and the following steps illustrate how it fails. The intended reading is the interpretation which would not violate Hurford's Constraint, and the meaning is syntactically represented as in (30). Unlike the analyses of Singh (2008) and Fox and Spector (2018), the current proposal does not prohibit the strengthening in the second disjunct.

(30) # Anna ate $[\text{all}]_F$ of the cookies or Anna ate $[\text{O}_{ALT}\ \text{some}]_F$ of the cookies.

The focus value of $[\![[\text{all}]_F[\!]]\!]_f$ is a set of quantificational determiner meanings, and ALT_{all} contains only those quantifiers whose structural complexities do not exceed that of *all*. Among those, only a couple of simplex determiners are inconsistent with *all*: *no* and possibly *few*. Importantly, the mutually exclusive ALT_{all} does not include either the logical meaning of *some* or the strengthened meaning of *some*. The former is not mutually exclusive with *all*, and the latter involves more structure with the additional ingredient, namely O_{ALT}. In contrast with the previous case with *some*, the mutually exclusive set cannot be expanded any further. Strengthening the meaning of *all* is not possible, as it is the strongest among the quantifiers that have entailment relations with *all*. The failure of strengthening is the cause of the unchanged ALT_{all}.

There is more than one way to cash out this intuition. The critical difference depends on the assumption on what primarily guides the expansion of ALT_α, whether it is the syntactic structure of α or the semantic denotation of α. If the structure is the key factor, then, the structure [O_{ALT} *all*] must be blocked. Otherwise, the inclusion of other quantifiers of the form [O_{ALT} *Q*], such as [O_{ALT} *some*], would be allowed, and we would incorrectly predict (30) to be legitimate. The ban on semantically vacuous attachment of O_{ALT} can be facilitated by following Fox's (2000) notion of derivational economy[12]. If the primary force of expansion of ALT_α is the semantic denotation of α, on the other hand, no special reference to the syntax of O_{ALT} is needed. Since the ordinary semantic value of [O_{ALT} *all*] is identical to that of *all*, ALT_{all} is unchanged. Therefore, the set cannot include the ordinary value of [O_{ALT} *some*], and *all* remains inadequate as a contrast antecedent for [O_{ALT} *some*][13].

[12]As mentioned in connection with example (10), the attachment of O_{ALT} in the acceptable structure, [O_{ALT}*some*] … or *all* …], is globally vacuous although it is meaningful within the first disjunct. It must be the case, therefore, that the ban on semantically vacuous attachment of O_{ALT} is computed locally. I am grateful to an anonymous reviewer for drawing my attention to this point.

[13]In the original formulation of Katzir's (2007) account, a structurally more complex alternative is added when it is subsequently mentioned explicitly. The main motivation for such a move is based on Matsumoto's (1995) example.

(i) It was warm yesterday, and it is a little bit more than warm today (Matsumoto, 1995, ex. 39, p. 44).

The first sentence seems to implicate that it was not a little more than warm yesterday, and to generate the implicature, "a little bit more than warm" must be a (stronger) scalar alternative that must be considered for "warm" even though it is obviously more complex than "warm." This mechanism may be needed for the computation of implicatures, but it is incompatible with the current proposal, for which subsequent mention of scalar alternatives itself is the subject of regulation.

To tell the two approaches apart, we examine a case of a stronger but not the strongest scalar item as a contrast antecedent. Consider (31), which patterns together with the infelicitous *all–some* order[14].

(31) a. #Anna ate most of the cookies, or Anna ate some of the cookies.

b. Anna ate [O_{ALT} most] of the cookies, or Anna ate [O_{ALT} some] of the cookies.

Unlike *all*, the attachment of O_{ALT} to *most* is not vacuous. Therefore, the structure will not be ruled out by the relevant Economy consideration. Nonetheless, it cannot serve as a good contrast antecedent for *some*, or [O_{ALT} *some*] to be more precise. If the structure [O_{ALT} *most*] would license the inclusion of all quantifiers of the form of [O_{ALT} *Q*], provided that they are mutually exclusive with [O_{ALT} *most*], (31a) is predicted to be well-formed, contrary to fact. We can conclude from (31) that Katzir's structural-complexity-based condition does not determine the membership of alternatives although it can serve as a restriction on it. In other words, the alternatives in ALT_α must be structurally as complex as or less complex than α, but not all candidates that satisfy that structural requirement can be in ALT_α even if the mutual exclusiveness condition is met.

When α is strengthened via O_{ALT}, the ordinary semantic value of [O_{ALT} α] drives the expansion of $\text{ALT}_{O[\alpha]}$. More concretely, we suggest that the exclusion of stronger alternatives makes those very alternatives highly relevant and be included in $\text{ALT}_{O[\alpha]}$:

(32) a. When [O_{ALT} α] is not vacuous, there is β, a stronger alternative to α, that is to be negated by [O_{ALT} α].

b. This process makes the issue of whether β or not β highly relevant, and β is added to the mutually exclusive ALT_α.

In example (31), the relevant stronger alternative is *all*. Thus, ALT_{most} now includes *all* but not *some* (or more precisely *some but not all/most*). Therefore, (31) remains infelicitous. With *all*, there is no stronger alternative to be excluded, and as a consequence, ALT_{all} does not expand. While we may still wish to ban the semantically vacuous attachment of O_{ALT} for an independent reason, such a prohibition is not necessary to account for the ordering asymmetry.

To sum up, the condition on contrast antecedents can provide a straightforward account for the ordering asymmetry. The examination has so far focused on the *some* – *all* pair in disjunction, but the analysis is generalizable to other scalar items in contrastive environments in general.

(33) **Successful Weak–Strong Order**: For a pair of scalar items α, β, where α is weaker/less informative than β and α linearly precedes β,

a. The mutually exclusive set of alternatives for $[\![\alpha]\!]$ cannot include $[\![\beta]\!]$ because they are not mutually exclusive.

[14]I am grateful to an anonymous reviewer for pointing out this example to me.

b. However, α can be strengthened via $[O_{ALT}\ [\alpha]]$, and $[O_{ALT}\ [\alpha]]$ now excludes $[\![\beta]\!]$.

c. This exclusion process makes $[\![\beta]\!]$ highly relevant, and it is now included in the mutually exclusive set of alternatives for $[\![[O_{ALT}[\alpha]]]\!]$.

d. Therefore, the order of α – β in contrastive environments is felicitous, as α, or more strictly speaking, $[O_{ALT}\ [\alpha]]$ is a good contrast antecedent.

(34) **Unsuccessful Strong–Weak Order**: For a pair of scalar items α, β, where α is stronger/more informative than β and α linearly precedes β,

a. The mutually exclusive set of alternatives for $[\![\alpha]\!]$ can include neither $[\![\beta]\!]$ nor $[\![[O_{ALT}[\beta]]]\!]$. The former is not mutually exclusive with $[\![\alpha]\!]$, and the latter is structurally more complex than $[\![\alpha]\!]$.

b. Since α is the stronger scalar item, the strengthening via $[O_{ALT}\ [\alpha]]$ does not exclude $[\![\beta]\!]$. Thus, the mutually exclusive set of alternatives still cannot contain $[\![\beta]\!]$ or $[\![[O_{ALT}[\beta]]]\!]$.

c. Therefore, the order of α – β in contrastive environments is infelicitous, as α is not a good contrast antecedent for β or $[O_{ALT}\ [\beta]]$.

The critical difference between (33) and (34) is whether a contrast antecedent can be strengthened in such a way that its contrast mate is regarded highly relevant. When the weaker item is strengthened, the stronger item becomes highly relevant and is included in the set of mutually exclusive alternatives while the stronger item does not undergo a comparable process. This difference leads to the ordering asymmetry. It is also highly important to note that the proposed explanation requires the ordinary value of a contrasted scalar item to be strengthened, as in (33b). In the disjunction examples that we examined in this section, for instance, the strengthening must take place locally within the first disjunct. The grammatical approach to scalar implicatures makes an easy choice for this process. While the current proposal departs from Singh's (2008) and Fox and Spector's (2018) grammar-based accounts of the ordering asymmetry, it still endorses the grammaticist's approach to scalar implicatures, as these previous analyses do.

4.3. Beyond the Basic Asymmetry

We have so far analyzed clear cases of Singh's ordering asymmetry in broader contrastive environments. However, there are several more complicated and subtle issues that go beyond the basic pattern, and this section examines whether these challenges are met by the proposed account.

First of all, observe that polar opposite contrasts are felicitous in either order. This is predicted as either scalar item can have the other as a mutually exclusive alternative.

(35) a. Anna ate all of the cookies or none of the cookies.

b. Anna ate none of the cookies or all of the cookies.

(36) a. Anna is an excellent cook or a bad cook

b. Anna is a bad book or an excellent cook.

Interestingly, the order of *strong – weak* becomes acceptable if the polar opposite of the strong item is also mentioned explicitly as an alternative, as shown below.

(37) a. It is not known whether Anna ate all, (or) some or none of the cookies.

b. (Of course,) Anna ate all, (or) some or none of the cookies. We just don't know which is true.

(38) a. We are debating whether Anna is an excellent cook, a good cook or a bad cook.

b. This spa has many choices. A hot bath, a warm bath, and a cold bath. A dry sauna and a steam room, too. Enjoy!

In these examples, the weaker items in the middle are strengthened: *some (but not all)* in (37), *good (but not excellent)* in (38a) and *warm (but not hot)* in (38b). Intuitively speaking, the mention of polar opposite alternatives can create a "multiple-choice-survey" like context where middle categories are carved out as independent categories. A theoretical explanation of this "carving out the middle" effect is not straightforward. If the middle scalar items are strengthened with O_{ALT} to be mutually exclusive with the stronger items, the question is how they are legitimate alternatives even though they are more structurally complex alternatives. It is noteworthy, in connection to this puzzle, that an overtly strengthened expression can seem to be an alternative in a similar context. Consider *excellent* and *very good*, for instance. Clearly, *very good* should not be a good focus alternative to *excellent*, as the former is more complex with an overt intensifier. As expected, the disjunctive structure, *excellent or very good*, sounds quite infelicitous. The structure improves dramatically, however, when it is embedded in a list like *excellent, very good, good, satisfactory, or bad*, and *very good* in this context is indeed interpreted as 'very good but not excellent'. At this point, the explanation remains rather descriptive, but the effects are robust.

The improvement by polar contrast is the strongest if the relevant polar alternatives are mentioned in a "list-like" fashion, as in the examples above. The following are examples where the polar alternatives are mentioned separately from the relevant disjunctions.

(39) a. ?We know it is impossible that Anna ate none of the cookies, as she simply cannot resist cookies. The question is whether she ate all of them or some of them.

b. ?We all agree that Anna is definitely not a bad student. We are still debating, however, whether Anna is an excellent student or a good student.

While they do not sound as natural as (37) and (38), many speakers find them more or less acceptable. Does the presence of a polar alternative need to be overt, or does it suffice if the utterance context clearly indicates its existence? Here are some test cases.

(40) Context: Professor Smith is telling her students how they performed in the recent exam in very general terms. Not surprisingly, no students got zero points. Also not surprisingly, not all students got all the answers right. And one of the students is saying to himself:

?(?)I wonder if I got all the answers right or some of them right…

(41) Two professors are evaluating a qualifying paper by one of their graduate students. They have already agreed that the paper should pass. They are now debating whether the paper should get "pass" or "pass with distinction," the higher honor. One of them says:

?(?) The question is whether it is an excellent paper or a good paper.

The judgment of these examples is delicate. They are perhaps a little odder than (39), and the addition of *only* or *just* to the second disjuncts makes them more natural. Nonetheless, the sentences sound much better in these contexts than when they are presented without any specific background information. In (40), the possibility of someone getting no answers right was considered (and eliminated). (41) indicates that the question of whether the paper is a bad paper was relevant prior the utterance context. Contextual information of this kind still encourages the hearer to maintain *no* and *bad* as alternatives in the context, and this appears to lead to the inclusion of the strengthened version of *some* and *good* in the set of mutually exclusive alternatives.

The gradual decline of improvement from (37)/(38) to (39) to (40)/(41) is very indicative of a more layered and nuanced situation than what the categorical labeling of felicity implies. The generalization is that the level of salience of polar alternatives corresponds to the degree of ease of making the middle ground alternative available. When one makes a list of choices in the form of disjunction or conjunction, all the to-be-mentioned alternatives are highly salient at the time of utterance. Without overt mention, the salience of the relevant polar alternative decreases, but the context can still sustain it to a certain level so that one may find a way to access the middle ground.

The discussion leading up to this point raises a new question: In general, how bad or how infelicitous is the "strong-weak" order of scalar items in the first place? So far I have followed the practice adopted in the previous studies on this topic and used categorical labels. A sentence in question is either felicitous or infelicitous (marked by #). There is definitely something stable, namely the preference of the "weak-strong" order in contrastive environments. However, the overall judgment patterns of the native speakers that I have consulted for this project were not categorical. The supposedly unacceptable "strong – weak" order is awkward but not hopelessly infelicitous, and the intended contrast is somehow achievable with some effort. Fox and Spector's (2018) corpus study is consistent with this overall judgment pattern. For instance, they found in the Corpus of Contemporary American English (http://corpus.byu.edu/coca/) 53 instances of *all – some*, as opposed to 396 of *some – all*, in disjunction. While these statistics confirm the strong preference of the "weak – strong" order of two scalar items in disjunction, they also suggest that the other less preferred order can occur with some regularity. Our findings regarding the role of polar opposite alternatives suggest that the relatively mild infelicity of the "strong-weak" order of scalar items can be due to our willingness to imagine the weaker item to be the relevant middle ground even when there is no clear indication of the presence of the polar opposite to the stronger item. It is definitely a kind of process that requires some effort, contrasting sharply with the "weak – strong" order, which generates a suitable set of alternatives with ease by appealing to O_{ALT}, a readily available procedure.

While the ordering asymmetry is generally more gradable than categorical, there is an interesting case in which the ill-effects of the "strong – weak order" are much more pronounced, to the extent that the judgment seems categorical. It is a case of the contrast between "A" and "A and B." The two paradigms, one in disjunction and the other in a *but*-conjunction, are shown below.

(42) a. Andy insulted [ANNA] or [Anna AND her SISTER].
b. # Andy insulted [Anna AND her SISTER] or [ANNA].
c. Andy insulted [Anna AND her SISTER] or only ANNA.

(43) a. ANDY insulted ANNA, but BILLY insulted Anna AND her SISTER.
b. #BILLY insulted ANNA and her SISTER, but ANDY insulted ANNA.
c. BILLY insulted ANNA and her SISTER, but ANDY insulted only ANNA.

The infelicity of (42b) and (43b) is very strong. These cases are judged far worse than the *all - some* or the other scalar items that we have examined so far. In order to generate the intended interpretations [= the interpretations comparable to (42a) and (43a)], it is necessary to add *only* or *just*. The question is why the effects are so strong in these examples. Additionally, the felicity of (42a) and (43a) presents an interesting puzzle. In these examples, the proper name *Anna* is strengthened to [O_{ALT} *Anna*] so that it means "only Anna." Since (42a) and (43a) are perfectly acceptable, it indicates, under the current proposal, that the strengthened *Anna* can be a contrast antecedent for *Anna and her sister*, which is, at least superficially, more structurally complex. Therefore, $ALT_{O[Anna]}$ should not include the meaning of *Anna and her sister*, which is an alternative to $ALT_{O[Anna]}$ under Katzir's definition of alternatives. One possible solution is to adopt Sauerland's (2004) artificial binary conjunctions L and R involved in these cases. With these conjunctions, an atomic formula can be regarded as structurally parallel to conjunction and disjunction. Semantically, L returns the semantic value of what comes to its left: for any ϕ, ψ, $[\![\phi\ L\ \psi]\!] = [\![\phi]\!]$. R is the mirror image of L: $[\![\phi\ R\ \psi]\!] = [\![\psi]\!]$. Sauerland's conjunctors make the proper name *Anna* as structurally complex as one with the conjoined counterpart, and the inclusion of *Anna and her sister* in the set of alternatives becomes possible. There may be other solutions with or without modifying Katzir's concept of structural complexity[15]. I assume that the challenge can be overcome and will focus on the issue of the strong infelicity of (42b) and (43b).

I suggest that this stronger ill-effect coincides with the failure of the "carving out the middle" strategy. As the following examples show, adding the polar opposite of "(both) A and B," namely "neither A nor B," helps to strengthen "(either) A or B" to "(either) A or B but not both," but it still fails to strengthen "A" to "only A" or "A but not B."

(44) a. Andy insulted [(both) Anna and her sister], (or) either one of them, or neither of them. I don't know exactly what happened.
b. # Andy insulted [(both) Anna and her sister], (or) Anna, (or) her sister, or neither of them. I don't know exactly what happened.
c. Andy insulted [(both) Anna and her sister], (or) only Anna, (or) only her sister, or neither of them. I don't know exactly what happened.

In (44b), neither "A" alone nor "B" alone can independently be carved out and strengthened. If the milder infelicity of other scalar pairs is derived from the "carving out the middle role" of a polar opposite alternative, as the current proposal hypothesizes, the strong infelicity of (42b) and (43b) is correctly predicted. As for why each conjunct fails to be strengthened even in the "list-like" context with the polar alternatives, I suggest that it is due to the non-unique nature of the scales involved with them. For all the other scalar contrasts, there are unique scales: "all – some – no," "excellent – good – bad," "hot – warm – cold," for example[16]. For conjunction, however, both conjuncts, or in Sauerland's (2004) *L* and *R*, cannot be placed on the same scale. Instead, there are two independent scales: "A and B – A – either A or B – neither A nor B" and "A and B – B – either A or B – neither A nor B." This parallel existence of two independent scales makes it not possible for "A" or "B" to be strengthened. If we eliminate A/B from the two scales, on the other hand, they merge into one scale: "A and B – either A or B – neither A nor B." Thus, *either A or B* can be made into an independent middle category, as the acceptable example (44c) shows.

Another question that naturally arises at this point is: does the addition of a non-polar-opposite alternative also help? One example related to this question is discussed in Fox and Spector (2018). They claim that previous mention of *most* makes the *all - some* order acceptable[17].

(45) A: Did John do most of the homework?
B: No. He did all of it or some of it. = Fox and Spector [2018, (47)]

While this example is presented as an acceptable case of the "all – some" order, I have not had much success in replicating as clear a judgment as Fox and Spector report. Some of my consultants felt that the sentence improves somewhat, but others didn't find it noticeably better. Instead of contradicting Fox and Spector's report, I assume that (45) manifests improvement which is nonetheless weaker and more variable than Fox and Spector's portrayal[18]. The first step toward a possible analysis is to treat A's question, "did John do most of the homework?", as a more specified version of the general question, "How much of the homework did John do?", which we may regard as Question-under-Discussion (QUD) in (45) (cf. Roberts, 1996)[19]. From then on, the conversation in (45) proceeds as follows.

(46) a. B's "no" response means that (B believes/knows that) it is not the case that John did most of the homework.
b. However, B's "no" response in (45) must be the negation of the strengthened meaning of *most*, namely "most but not all," because the negation of the unstrengthened *most* would be contradictory with B's subsequent statement, *he did all or some of it*.
c. Finally, B follows up and addresses the QUD, "how much of the homework did John do?", by offering a positive answer to one of the two other related polar questions, "Did John do all of the homework?", "Did John do some of the homework?".

The process in (46) creates a context in which there are several polar questions out of the QUD. They in turn can lead to a multiple-choice-answer scenario with respect to the QUD: *all, most, some* and *no*. As we discussed above, a multiple-choice context encourages the choices to be mutually exclusive, and the

[15] In this instance, the structural complexity may not be playing any role. According to the reasoning behind the expansion of alternatives in (32), [O_{ALT} *Anna*] excludes *(both) Anna and her sister*, which makes *(both) Anna and her sister* highly relevant and be included in $\text{ALT}_{O[Anna]}$. The felicity of (43a) suggests that this step is good enough, making the restriction based on structural complexity irrelevant.

[16] I am simplifying the discussion by ignoring monotonicity. In other words, I am putting positive scales (e.g., "excellent – good") and negative scales ("bad – terrible") together. While monotonicity may play an important role in generating implicatures [as argued by Horn (1989) and Matsumoto (1995)], I assume it is not relevant for focus alternatives.

[17] For this improvement effect, Fox and Spector (2018) use the notion of "distant entailing disjunction." Since their account assumes the asymmetry to be confined within disjunctive structure, it will not reviewed in this paper. See section 5.2.2.1 of Fox and Spector (2018) for more discussion.

[18] Fox and Spector (2018) acknowledge that some of the judgments they report are delicate. "As we develop our proposal we will be presenting a variety of very detailed predictions, some of which will involve rather subtle contrasts in acceptability judgments. We are not always as confident about these judgments as we would like to be. Nevertheless we think that stating the predictions explicitly would be useful in understanding the nature of our proposal. Within the current proposal, the effect can be derived via the following steps" (Fox and Spector, 2018, p. 7). (45) seems to be one such instance involving a subtle judgment.

[19] The concept of "more specified question" is not identical to the notion of *subquestion* discussed in Roberts (1996), Beck and Sharvit (2002), and Büring (2003), and others. An exhaustive answer to a subquestion of Q is a partial answer to Q. (i) is an example of the subquestion relation, in which the last two questions are subquestions of the first.

(i) Who recommended who? Who did Maria recommend? And who did Anna recommend?

A "more specified question" of Q is a polar question made out of one possible answer to Q, as illustrated in (ii). A positive answer to a more specified question answers the general question, but a negative answer does not.

(ii) How much was the ticket to the concert? Was it $100? $150?

ordering asymmetry of the scalar items often disappears. In Fox and Spector's example (45), the presence of the relevant QUD is only implied, but it can be made explicit, as in (47). Its addition seems to enhance the improvement effect.

(47) A: How much of the homework did John do? Did he do most of it?
B: No. He did all of it or some of it.

This kind of gradient improvement is expected, as it patterns with the variable improvement effects that we saw above in connection with other multiple-choice environments.

To summarize our discussion in this subsection, we have examined several cases in which the allegedly infelicitous "strong – weak" order becomes acceptable or is at least tolerated. The presence of a polar-opposite alternative is playing a pivotal role in increasing the felicity of this disadvantaged order. The salience of a relevant polar alternative can be raised in a variety of ways, and its impact on the ordering asymmetry surfaces as the gradable/variable judgment of acceptability.

4.4. Fine-Tuning the Contrast Antecedent Condition

So far, the target of the Contrast Antecedent Condition has been focused scalar items themselves. The key ingredient is that a good antecedent in a scalar contrast has the potential to be strengthened, and the way I have been describing the processes seems to imply that the required strengthening mechanism is lexical. While the phenomena examined in this paper are certainly compatible with the lexicalist version of the grammaticist approach, that is not the only option[20]. As reviewed earlier, the grammaticist approach of Chierchia et al. (2012), for instance, employs a clausal operator to achieve the required exhaustification. It is worth considering how the proposed condition can accommodate the sentential operator O_{ALT}.

(48) **Contrast Antecedent Condition'**: Alternative to (26)
For any phrase α and α' such that α is dominated by α', when α is contrastively focused, there must be β which precedes α and is dominated by β' which generates $ALT_{\beta'}$, a set of alternatives for β', such that
(i) it is a subset of the focus semantic value of β',
(ii) its members are mutually exclusive, and
(iii) it includes both the ordinary value of β' and that of α'.

Here is a case study of this definition.

(49) a. [O_{ALT} [Anna ate [some]$_F$ of the cookies]] or [Anna ate [all]$_F$ of the cookies]]
b. α = *all*
β = *some*
α' = *Anna ate all of the cookies*
β' = O_{ALT} *[Anna ate some of the cookies]*
c. $[\![\alpha']\!]_o$ = Anna ate all of the cookies.
$ALT_{\beta'}$ = {Anna ate some but not all of the cookies, Anna ate all of the cookies, Anna ate none of the cookies}
d. Therefore, $[\![\alpha']\!]_o \in ALT_{\beta'}$, and the condition is met.

The current version of the Contrast Antecedent Condition needs further fine-tuning in order to accommodate inter-speaker contrasts. In a dialog, a speaker can make a contrastive statement in relation to what has been uttered by another conversation partner. For this partner, there is no notion of "planning ahead" or "forecasting" for what kind of contrastive statement may follow her statement. When a speaker utters a contrastive statement with the linguistic expression α being focused, the contrast antecedent of α may or may not be focused. The following conversations exemplify this scenario. There are two noteworthy points in these examples: (i) it is possible to contrast a scalar item even when the corresponding scalar item in the previous sentence is not focused, and (ii) the ordering asymmetry is still present.

(50) A: What are the students in your department working on these days?
B: Let's see, some of ANNA's students are working on IMPLICATURE, and…
C: Not true! ALL of her students are working on implicature.

(51) A: What are the students in your department working on these days?
B: Let's see, all of ANNA's students are working on IMPLICATURE, and…
C: Not true! #(Only) SOME of her students are working on implicature.

(52) A: Which students in your program do you think highly of?
B: Let's see, ANNA is a good student, and so is BERTHA, and …
C: Wait, Bertha is a BRILLIANT student. She should not be mentioned in the same sentence with Anna.

(53) A: Which students in your program do you think highly of?
B: Let's see, ANNA is a brilliant student, and so is BERTHA, and …
C: Wait, Bertha is #(just) a GOOD student. She should not be mentioned in the same sentence with Anna.

[20]The data in this paper do not distinguish the approach that encodes the strengthened meaning at the lexical level (e.g., Levinson, 2000; Chierchia, 2004) from the clausal operator approach. As a matter of fact, a version of the lexical approach is put forth for contrasted scalar items by Geurts (2010, Chapter 8), who is a committed advocate of the globalist approach. He argues that a contrasted scalar item undergoes lexical narrowing/strengthening, which would be compatible with the current proposal. The choice between the lexicalist or the clausal operator approaches must come from facts other than the ones considered here. Sauerland (2012, 2014) presents a critical review of the lexicalist appraoch based on the observation that there are numerous instances of intermediate implicatures that the lexicalist approach predicts to be impossible. Tomioka (2019) endorses Sauerland's conclusion by closely examining Geurts' hypothesis with special attention to the scope properties of contrasted scalar quantifiers. It is shown that the scope of a scalar implicature does not match that of the relevant scalar item, contrary to what Geurts' analysis predicts. All in all, the lexicalist approach faces more empirical challenges than the clausal operator approach.

In light of these examples, the condition must be modified. First, we need to define the concept of *potential focus values*.

(54) A potential focus value of a linguistic expression γ is the focus value of γ', which is identical to γ except for the location of F-marking.

We add this concept to (48).

(55) **Contrast Antecedent Condition"**: Modified version of (48)
For any phrase α and α' such that α is dominated by α', when α is contrastively focused, there must be β which precedes α and is dominated by β' which generates $ALT_{\beta'}$, a set of alternatives for β', such that
(i) it is a subset of the actual or a potential focus semantic value of β',
(ii) its members are mutually exclusive, and
(iii) it includes both the ordinary value of β' and that of α'.

Consider (50), for example. The constituent that corresponds to the contrast antecedent is *some of Anna's students are working on implicature*, where the main foci fall on *Anna's* and *implicature*. Thus, the actual focus value of this sentence is a set of propositions of the form "some of X's students are working on Y" where X is a professor and Y is a research topic. This would not be a good basis of the right contrast antecedent for the subsequent contrastive statement. There is a potential focus value of this sentence where the main focus shifts to *some*. Then, that hypothetical focus value is a proposition of the form "Q of Anna's students are working on implicatures," and after *some* is strengthened via O_{ALT}, it becomes a mutually exclusive set, namely {some but not all of Anna's students are working on implicatures, all of Anna's students are working on implicatures, none of Anna's students are working on implicatures}, which satisfies the condition.

4.5. No Asymmetry in Non-contrastive Environments

The current proposal predicts that the "stronger – weaker" order of scalar items causes no ill effects when the scalar items are not contrasted. This prediction is borne out.

(56) *Context: Your colleague is directing a study abroad program to Japan, and you are worried that the students may not have sufficient international experience to do well in the program. Your colleague says she is not worried, saying . . .*

All the students have been to Europe. Some of them have actually been to Japan before.

(57) *How was your hotel in Paris?*

Oh, the room was absolutely beautiful, the service was impeccable, and the location was good. So, we were very happy.

These examples are natural with the potentially problematic order of scalar items. The difference here is that they are not contrasted with each other. In both cases, the sentences collectively give an answer to the QUDs.

Note that the weaker scalar items are strengthened. The second sentence in (56) is most naturally interpreted as "some but not all," and in (57), the speaker must have meant that the hotel's location was good but not exceedingly so. What is remarkable here is that the addition of an overt exhaustive expression, such as *only* or *just*, is infelicitous. This is one of the instances in which the implicit exhaustification cannot be paraphrased by adding an overt exhaustive expression, and it will guide the discussion in the next section: Why does the presence of *only* or *just* save the otherwise infelicitous order of scalar items in contrast?

5. COMMENTS ON *ONLY*

One notable difference between the current proposal and the previous analyses is the view on the distribution of O_{ALT}. In the account developed here, there is no restriction on it. It can appear, in principle, in the second disjunct in disjunction, provided that a good contrast antecedent is present in the preceding context. Both Singh (2008) and Fox and Spector (2018) attribute the ordering asymmetry to a certain distributional restriction on O_{ALT}, and their approach might come from the following train of thought: (i) O_{ALT} is the silent version of *only*, (ii) "all or [O_{ALT} some]" is not good, but "all or only some" is good, (iii) thus, the constraint is about how to regulate O_{ALT}.

Whether these authors indeed had this line of reasoning or not, the issue of *only* is unavoidable for the contrast based account as well. How does the presence of *only* or *just* make the otherwise unnatural order acceptable? My answer to this question begins with the objection to the practice of calling O_{ALT} the silent version of *only*. First of all, there are quantifiers that elicit the "not all" interpretations but resist the attachment of *only*, as discussed in Al Khatib (2013).

(58) *Why are the students so upset?*
  a. Both Professor Suzuki and Professor Tanaka flunked many/most of the students. $\rightsquigarrow$ Both professors are such that they did not flunk all of the students.
  b. # Both Professor Suzuki and Professor Tanaka only passed MANY/MOST of the students.

(58a) can seem to generate the local implicature that is paraphrased after $\rightsquigarrow$, which means, under the grammaticist system, that O_{ALT} is inserted below the subject. The operator successfully generates the "not all" implicature. On the other hand, the overt insertion of *only* is infelicitous in this context[21]. This means that O_{ALT} can exhaustify when *only* cannot.

It should also be noted that *only some* often adds something other than the not-all meaning.

[21] The two anonymous reviewers pointed out that *only most* is acceptable when it is explicitly contrasted with *all*. In the example above, however, there is no such contrast, but *most* still generates the relevant scalar implicature.

(59) *Did your relatives come to your wedding?*
A: Some of them did.
A': Some but not all of them did.
A'': Only some of them did.

The answer A is functionally equivalent to A', as we most naturally understand the meaning of *some* being strengthened. The equivalence between A and A'' is much less clear. A'' definitely indicates that not all of the relatives came, but it tends to communicate something extra, such as a sense of disappointment. We can highlight the difference by adding "I thought that none of them showed up." The addition is felicitous with Response A, but it is distinctly odd with Response A'', as demonstrated below.

(60) *Did your relatives come to your wedding?*
A: Some of them did. I thought that none them would show up.
A'': Only some of them did. #I thought that none of them would show up.

This extra meaning is related to the scalar use of *only*, which was noted by Horn (1969). The following sentences exemplify the scalar meaning of *only*.

(61) a. This time, Usain Bolt only got the bronze medal in the men's 100 m race.
b. My friend Joshua and I were short-listed for the same job at the company. Joshua had a meeting with the CEO, but I was interviewed only by the vice president.

We all know that for any given category of competition, one individual can receive no more than one medal. Thus, the semantic contribution of *only* in (61a) is not exhaustivity/exclusivity. Rather, it adds the meaning that what Usain Bolt got this time is a lesser medal than before. Similarly, (61b) can be uttered truthfully even when the speaker was also interviewed by some people other than the vice president. In these instances, *only* does not negate non-weaker alternatives. It instead indicates that the said content is "lower in the relevant scale" than some standard that was expected or was made salient in the context[22].

Turning our attention back to scalar items, it is clear that "*only* + a scalar item" generates the scalar meaning, not the exhaustivity meaning. What happens when "*only* + a scalar item" is used in a contrastive environment?

(62) a. Anna ate all of the cookies or only some of the cookies.
b. Anna read all of the textbooks, but Bertha read only some of them.
c. While Anna read all of the textbooks, Bertha read only some of them.

In these examples, *only some* does not demand the presence of a contrast antecedent that provides a mutual exclusive set. Rather, it requires an antecedent that provides the standard of comparison with which the prejacent of *only* is evaluated. Thus, the meaning of (62a) is paraphrased as Anna ate all of the cookies or she ate some of the cookies, which is lower than some standard in the relevant scale. The relevant scale is most naturally understood to be proportions of the cookies consumed, and the salient standard is eating all the cookies, which the first disjunct provides.

The same strategy works for other scale-mate pairs with *only* or *just*.

(63) a. Anna's sister is brilliant, but Anna herself is just smart.
b. Anna's sister is brilliant, but Anna herself is smart, which is lower in the relevant scale than being brilliant. ⇝ Anna is smart but not brilliant.

(64) a. When you ride a motorcycle, wearing a helmet is required in Japan. But in the U.S., it is only recommended.
b. … But in the U.S., it is only recommended, which is lower in the relevant scale than being required. ⇝ In the U.S., it is recommended but not required.

To sum up, calling O_{ALT} the silent version of *only* is more misleading than useful since the overt *only* is not the kind of exhaustivity operator comparable to O_{ALT} when it combines with a scalar item. As a consequence, it is not surprising that the distributional pattern of *only* does not exactly match that of O_{ALT}. In particular, the addition of *only* improves the less-than-perfect order of two scalar items, and it is due to a different requirement imposed on the contrast antecedent for *only*.

6. FINAL THOUGHTS

This paper presents a novel analysis of Singh's (2008) paradigm of the ordering restriction on scalar items and its interaction with Hurford's Constraint. It is based on the entirely new generalization that the relevant restriction is found not only in disjunction but in contrastive environments in general. I argue that the source of the restriction is rooted in the inherent ordering asymmetry in making contrasts. One important ingredient of the analysis is a mutually exclusive set of alternatives. This concept has been proposed in the past, but its use is extended to regulate the contrastiveness generated by focus. It should also be noted that the proposed account for the ordering asymmetry is applicable to Hurford's Constraint itself, as disjunction is one of the many linguistic environments that evoke the sense of contrastiveness. Singh's Inconsistency Condition on disjuncts is a direct consequence of the mutual exclusiveness requirement on the set of alternatives in contrastive contexts. Another noteworthy aspect of the proposal is that it makes it necessary

[22] Schwarzschild (1996) attempts to provide a uniform analysis of the exhaustive and the scalar uses of *only* by using the scalar meaning as its base. Similarly, Zeevat (2009) proposes a "mirative" analysis of focus sensitive adverbs, in which the sole semantic contribution of *only* is the scalar meaning of "less than expected." The alleged exhaustivity meaning in a sentence with *only* is derivable with focus; even without *only*, the sentence has the exhaustive meaning, as it is typically considered as the complete exhaustive answer to a QUD.

that the scalar meaning, traditionally labeled as a conversational implicature, is grammatically generated so that it can become the basis of computing alternatives.

In the current proposal, disjunction is one sub-type of contrastive environment, and the previous analyses based on the syntax and semantics of disjunction could be considered dispensable. It is predicted, therefore, that there are no disjunction-specific facts in the domain of the ordering restriction. In this regard, there are a few cases, noted by Fox and Spector (2018), that challenge this prediction. According to Fox and Spector, the strong-weak order becomes felicitous when it is under the scope of a universal quantifier, as exemplified in the following minimal pair.

(65) Fox and Spector (52ab)
 a. #Either John did both the reading and the homework or he did the reading or the homework.
 b. Either everyone did both the reading and the homework or everyone did the reading or the homework.

I have not succeeded in eliciting solid and consistent judgments on this minimal pair from the native speakers I consulted. If the contrast is as clear as reported in Fox and Spector (2018), it is indeed quite puzzling not only because my proposal has no account to offer but also because it does not seem to carry over to non-disjunctive contrastive environments. For instance, it is hard to find the same kind of contrast in the following pair[23].

(66) A: Eric did both the reading and the homework.
 B: (#) Not true! He did the reading or the homework.

(67) A: Every student did both the reading and the homework.
 B: (#) Not true! Every student did the reading or the homework.

The presence of contrast in (65) and the lack thereof in (66)/(67) would be an unwelcome combination for the current proposal.

Fox and Spector (2018) further note that the *all–some* order becomes acceptable if the disjunction structure that contains the quantifiers is placed under the scope of such operators as a universal quantifier and an intensional operator.

(68) a. Every student solved all or some of the problems. = Fox and Spector [2018, (58b)]
 b. A new Harris Poll finds a plurality of Americans want all or most abortions to be illegal. = Fox and Spector [2018, (60)]

While these examples pose challenges to the current proposal, they are more complex than what they appear to be. Consider the imperative structure in (69), inspired by Fox and Spector's (2018) example (61):

(69) Please tell me all or some of the names of the suspects.

[23]In general, the *and-or* order is judged a little better than the *all-some* order [e.g., (20)], hence # is parenthesized. The point here is that there is no discernible contrast between the proper name and the quantified subjects.

As predicted by Fox and Spector, the *all–some* order is acceptable in this example. It seems, however, that the meaning of this sentence does not correspond to the one expected by their analysis. This sentence is more appropriately paraphrased as (70a), rather than (70b) and (70c), which suggests that what is involved in this disjunction is not the straightforward application of O_{ALT}.

(70) a. Please tell me all or at least some of the names of the suspects.
 b. Please tell me all or only some of the names of the suspects.
 c. Please tell me all or some but not all of the names of the suspects.

The most natural interpretation of (69) indicates that the first disjunct is the preferred choice, but if that cannot be achieved, the second option is still acceptable. Closer inspection reveals that this "concessive-like" meaning is also relevant in the abortion example (68b). The sentence can depict the following situation: some of those Americans want all abortions to be illegal, but others, while they want most abortions to be illegal, have some conflicted feelings about a small number of exceptional cases. For those people, it is too strong to say that they want most but not all abortions to be illegal, as they cannot decide whether the exceptional cases should be legal or illegal. A proper analysis of this type of disjunction should begin with the assumption that it is not a part of the generalization based on Hurford's Constraint since the meaning, which is paraphrased as "all or at least some," seems to violate the constraint in the first place. Incidentally, non-disjunctive environments also allow the *all–some* order if a similar, "concessive-like" relation is expressed.

(71) a. If you cannot tell me all the names of the suspects, please tell me some of them.
 b. A: Are you asking me to tell you all the names of the suspects? You know that I can't.
 B: OK, then, can you tell me some of their names?

If these concessive-like cases are set aside as independent problems, there remain only a small set of disjunction-specific puzzles, such as (65) and (68a), that go beyond the contrast-based analysis that I advocate in this paper. They involve rather subtle and variable judgments, however. Closer and more rigorous examination of the data is needed for the current proposal to move forward, and experimental research will be particularly welcome.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article and the supplementary material.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Al Khatib, S. (2013). *'Only' and association with negative antonyms* (Ph.D. thesis). Cambridge, MA: MIT.

Asher, N., and Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Beck, S., and Sharvit, Y. (2002). Pluralities of questions. *J. Seman*. 19, 105–157. doi: 10.1093/jos/19.2.105

Blakemore, D., and Carston, R. (2005). The pragmatics of sentential coordination with and. *Lingua* 115, 569–589. doi: 10.1016/j.lingua.2003.09.016

Büring, D. (1997). The great scope inversion conspiracy. *Linguist. Philos*. 20, 175–194. doi: 10.1023/A:1005397026866

Büring, D. (2003). On d-trees, beans, and b-accents. *Linguist. Philos*. 26, 511–545. doi: 10.1023/A:1025887707652

Chierchia, G. (2004). "Scalar implicatures, polarity phenomena and the syntax/pragmatics interface," in *Structures and Beyond: The Cartography of Syntactic Structure, Vol. 3*, ed A. Belletti (Oxford: Oxford University Press), 39–103.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguist. Inq*. 37, 535–590. doi: 10.1162/ling.2006.37.4.535

Chierchia, G., Fox, D., and Spector, B. (2012). "The grammatical view of scalar implicatures and the relationship between semantics and pragmatics," in *Semantics: An International Handbook of Natural Language Meaning*, eds K. von Heusinger, C. Maienborn, and P. Portner (Berlin: de Gruyter).

Fox, D. (2000). *Economy and Semantic Interpretation*. Cambridge, MA: MIT Press and MIT Working Papers in Linguistics.

Fox, D. (2007). "Free choice and the theory of scalar implicatures," in *Presupposition and Implicature in Compositional Semantics*, eds U. Sauerland and P. Stateva (Basingstoke: Palgrave Macmillan), 71–120. doi: 10.1057/9780230210752_4

Fox, D., and Katzir, R. (2011). On the characterization of alternatives. *Nat. Lang. Seman*. 19, 87–107. doi: 10.1007/s11050-010-9065-3

Fox, D., and Spector, B. (2018). Economy and embedded exhaustification. *Nat. Lang. Seman*. 26, 1–50. doi: 10.1007/s11050-017-9139-6

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York, NY: Academic Press.

Geurts, B. (2010). *Quantity Implicatures*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511975158

Horn, L. (1969). "A presuppositional analysis of *only* and *even*," in *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society* (Chicago, IL), 97–108.

Horn, L. (1972). *On the semantic properties of the logical operators in English* (Ph.D. thesis). Los Angeles, CA: UCLA.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago, IL; London: The University of Chicago Press.

Hurford, J. (1974). Exclusive or inclusive disjunction. *Found. Lang*. 11, 409–411.

Katzir, R. (2007). Structurally-defined alternatives. *Linguist. Philos*. 30, 669–690. doi: 10.1007/s10988-008-9029-y

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.

Lee, C. (2007). "Contrastive (predicate) topic, intonation, and scalar meanings," in *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, eds C. Lee and M. Gordon (Berlin: Springer Science & Business Media), 151–175. doi: 10.1007/978-1-4020-4796-1_9

Levinson, S. C. (2000). *Presumptive Meanings*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/5526.001.0001

Matsumoto, Y. (1995). The conversational condition on horn scales. *Linguist. Philos*. 18, 21–60. doi: 10.1007/BF00984960

Menéndez-Benito, P. (2010). On universal free choice items. *Nat. Lang. Seman*. 18, 33–64. doi: 10.1007/s11050-009-9050-x

Roberts, C. (1996). "Information structure in discourse: towards an integrated formal theory of pragmatics," *Ohio State University Working Papers in Linguistics*, eds J. H. Yoon and A. Kathol (Columbus, OH), 91–136.

Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semant*. 1, 117–121. doi: 10.1007/BF02342617

Sæbø, K. J. (2003). "Presupposition and contrast: German aber as a topic particle," in *Proceedings of Sinn und Bedeutung*, 257–271.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27:343–377. doi: 10.1023/B:LING.0000023378.71748.db

Sauerland, U. (2012). The computation of scalar implicatures: pragmatic, lexical or grammatical? *Lang. Linguist. Compass* 6, 36–49. doi: 10.1002/lnc3.321

Sauerland, U. (2014). "Intermediate scalar implicatures," in *Pragmatics, Semantics and the Case of Scalar Implicatures*, ed S. P. Reda (Basingstoke: Palgrave Macmillan), 72–98. doi: 10.1057/9781137333285_4

Schwarzschild, R. (1996). *Givenness and optimal focus* (Ms.). New Brunswick, NJ: Rutgers University.

Singh, R. (2008). On the interpretation of disjunction: asymmetric, incremental, and eager for inconsistency. *Linguist. Philos*. 31, 245–260. doi: 10.1007/s10988-008-9038-x

Tomioka, S. (2010). "Contrastive topics operate on speech acts," in *Information Structure from Different Perspectives*, eds M. Zimmermann and F. Caroline (Oxford: OUP), 753–773.

Tomioka, S. (2019). The scope of contrasted scalar items and intermediate implicatures. *English Linguist*. 36, 29–47.

Umbach, C. (2005). Contrast and information structure: a focus-based analysis of but. *Linguistics* 43, 207–232. doi: 10.1515/ling.2005.43.1.207

Wagner, M. (2012). "Focus and givenness: a unified approach," in *Contrasts and Positions in Information Structure*, eds I. Kučerová and A. Neeleman (Cambridge: Cambridge University Press), 102–147. doi: 10.1017/CBO9780511740084.007

Winterstein, G. (2012). What but-sentences argue for: an argumentative analysis of but. *Lingua* 122, 1864–1885. doi: 10.1016/j.lingua.2012.09.014

Zeevat, H. (2009). "Only" as a mirative particle," in *Focus at the Syntax-Semantics Interface: Working Papers of the SFB 732*, Vol. 3, ed A. R. Onea (Stuttgart: Stuttgart University), 121–141.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership