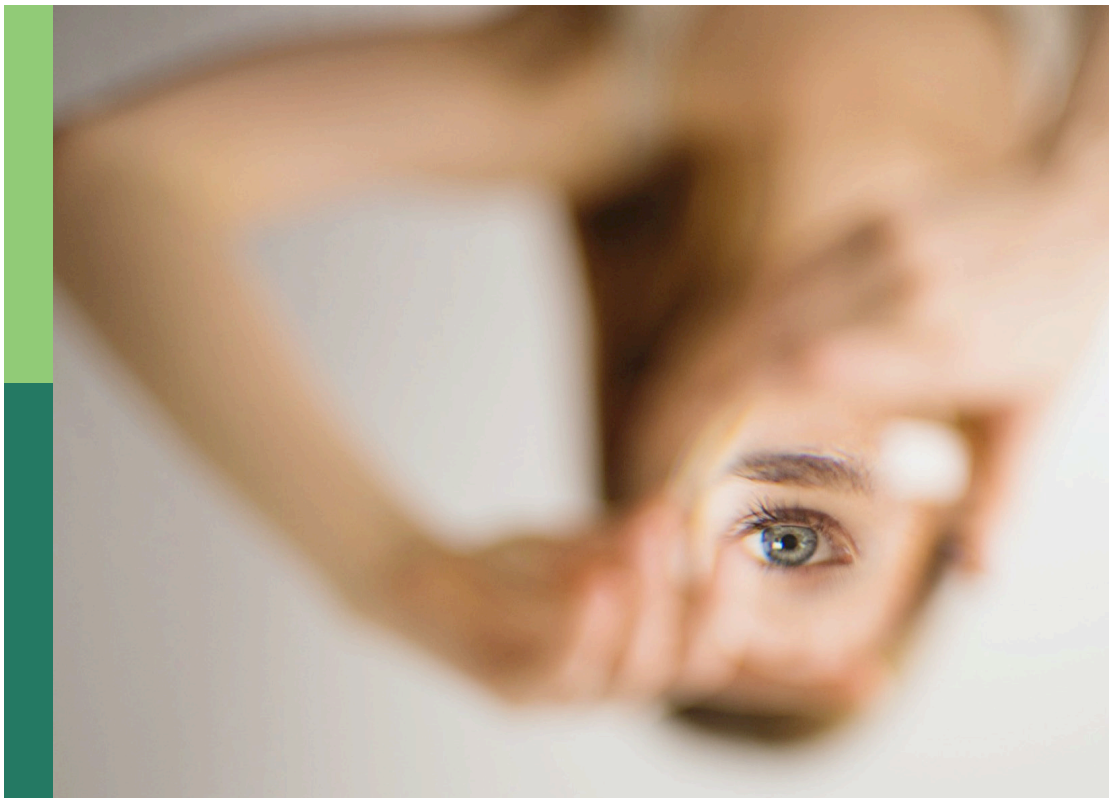
frontiers | Research Topics

# Relevance in mind

**Edited by**
Tim Wharton, Didier Maillat, Caroline Jagoe and Kate Scott

**Published in**
Frontiers in Psychology
Frontiers in Communication



ISSN 1664-8714
ISBN 978-2-8325-4134-0
DOI 10.3389/978-2-8325-4134-0

## 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

# Relevance in mind

**Topic editors**
Tim Wharton — University of Brighton, United Kingdom
Didier Maillat — Université de Fribourg, Switzerland
Caroline Jagoe — Trinity College Dublin, Ireland
Kate Scott — Kingston University, United Kingdom

**Citation**
Wharton, T., Maillat, D., Jagoe, C., Scott, K., eds. (2024). *Relevance in mind*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4134-0

# Table of contents

04 Editorial: Relevance in mind
Tim Wharton, Kate Scott, Didier Maillat and Caroline Jagoe

06 Communication and deniability: Moral and epistemic reactions to denials
Francesca Bonalumi, Feride Belma Bumin, Thom Scott-Phillips and Christophe Heintz

18 Social cognition and Relevance: How stereotypes impact the processing of definite and indefinite descriptions
Magali A. Mari and Misha-Laura Müller

34 Suppression of literal meaning in single and extended metaphors
Camilo R. Ronderos and Ingrid Lossius Falkum

44 Intonational production as a window into children's early pragmatic competence: The case of the Norwegian polarity focus and two *jo* particles
Line Sjøtun Helganger and Ingrid Lossius Falkum

58 Nutritional labeling, communication design, and relevance
Kate Scott

68 Strength is relevant: experimental evidence of strength as a marker of commitment
Kira Boulat and Didier Maillat

80 The relevance of words and the language/communication divide
Robyn Carston

95 Taking stock of an idiom's background assumptions: an alternative relevance theoretic account
Ira A. Noveck, Nicholas Griffen and Diana Mazzarella

107 Relevance and multimodal prosody: implications for L2 teaching and learning
Pauline Madella

115 Relevance theory and the social realities of communication
Marilynn Johnson

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 30 November 2023
DOI 10.3389/fpsyg.2023.1338336



OPEN ACCESS

EDITED AND REVIEWED BY
Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE
Tim Wharton
✉ t.wharton@brighton.ac.uk

RECEIVED 14 November 2023
ACCEPTED 16 November 2023
PUBLISHED 30 November 2023

CITATION
Wharton T, Scott K, Maillat D and Jagoe C (2023) Editorial: Relevance in mind. *Front. Psychol.* 14:1338336. doi: 10.3389/fpsyg.2023.1338336



# Editorial: Relevance in mind

Tim Wharton[1]*, Kate Scott[2], Didier Maillat[3] and Caroline Jagoe[4]

[1]School of Humanities and Social Science, University of Brighton, Brighton, United Kingdom, [2]School of Design, Kingston University, Kingston upon Thames, United Kingdom, [3]Department of English, Université de Fribourg, Fribourg, Switzerland, [4]School of Linguistic, Speech and Communication Science, Trinity College Dublin, Dublin, Ireland

KEYWORDS

**relevance, pragmatics, relevance theory, cognition, communication**

Editorial on the Research Topic
Relevance in mind

This *Frontiers* Research Topic is an attempt to push the envelope of relevance theory with particular attention to its implications for psychology and cognitive science and to disseminate the theory more widely, encouraging others to engage with the theory and better understand its capacity to broaden and deepen our understanding of all aspects of communication and cognition. With that in mind, the contributions to "*Relevance in Mind*" address one of the following three themes:

(1) Relevance theory in psychology and cognitive science
(2) Relevance theory, social communication, and social cognition
(3) Relevance theory: extending the boundaries from within

The papers are grouped according to the theme they most closely fit, while recognizing that these themes are not mutually exclusive.

## Theme 1: Relevance theory in psychology and cognitive science

In *Intonational production as a window into children's early pragmatic competence*, Helganger and Falkum investigate what the production of the Norwegian "Polarity Focus" intonation pattern by preschool children reveals about their early pragmatic development. The mastery of this pattern, they claim, can be seen as an early linguistic manifestation of relevance-driven cognitive abilities for the attribution of thoughts and epistemic vigilance toward propositional content. Noveck et al.'s *Taking stock of an idiom's background assumptions* argues that since relevance theory has tended to focus on the interpretation of metaphor and irony, there is a great deal of work to be done on the interpretation of idioms. They argue for a relevance theory approach in which idioms are explained through the fact that they activate presuppositional information. The new approach is confirmed through a pilot experiment.

In *Strength is relevant: experimental evidence of strength as a marker of commitment*, Boulat and Maillat explore the notion of "strength", one of the relevance-theoretic organizing principles responsible for ordering contextual assumptions. They argue for a theoretical notion in which strength is regarded as a marker of commitment, and—more generally—of the epistemic value of an utterance. This claim is supported through a set of new experiments in which levels of strength are manipulated and, in turn, shown to correlate with accuracy in a recollection task. Their results support their model and its implications are discussed.

Ronderos and Falkum's *Suppression of literal meaning in single and extended metaphors* tests recent theoretical claims made by Carston on the differences between the processing of single and extended metaphors. Their work builds on claims that processing single metaphors involves suppressing features related exclusively to the literal meaning. Their goal is to investigate whether suppression is also involved in the comprehension of extended metaphors, or whether—as Carston suggests—the literal meaning "lingers", thereby leading to the continued activation of such features. They suggest their results lend support for Carston's view.

## Theme 2: Relevance theory, social communication, and social cognition

Mari and Müller's paper *Social cognition and relevance* explores the impact of social cognition on the processing of linguistic information, demonstrating how gender and nationality-related stereotypes guide the relevance-based processing of definite and indefinite descriptions. Results show that information contradicting nationality stereotypes costs significantly more in terms of processing effort than information confirming stereotypes. Overall, the findings are consistent not only with research on stereotypes, but also the relevance theory claims on the relationship between effort and effects. In *Relevance theory and the social realities of communication*, Johnson considers one of the central tenets of intention-based theories of pragmatics: that the mental states of our interlocutors are altered on the basis of their recognition of our communicative intentions. She argues that this is not equally the case for all interlocutors and that according to various social factors, some bear an additional burden. By demonstrating how social factors affect the reality of the way social beings interact and communicate Johnson builds theoretical bridges between relevance theory and Fricker's work on *testimonial injustice*.

Bonalumi et al.'s *Communication and deniability: Moral and epistemic reactions to denials* looks at the potential effects of situations in which a speaker denies having meant what an audience understands them to have meant. They present experiments which explore those incentives a speaker might have to mislead their audience and the impact a speaker's denial might have on an audience's moral and epistemic assessments of what has been said. On the basis of their initial findings, they present an original analysis of how audiences react to denials which draws on the relevance theory approach to communication.

## Theme 3: Relevance theory: extending the boundaries from within

Carston's *The relevance of words and the language/communication divide* explores the idea that relevance theorists have tended to emphasize the communicative dimension of words (the construction of *ad hoc* senses, for example) at the expense of the morpho-syntactic side of language. Words, after all, are not only the building blocks of communicative exchanges. They are also the building blocks of linguistic form. The discussion suggests how the communicative side to words might interface with the computational (linguistic) one and how words effectively "straddle" the divide. It also presents evidence from populations with atypical development showing that both sides of the divide are affected differentially which suggests it is a natural one in human cognitive architecture.

Madella's *Relevance and multimodal prosody* presents the implications of analyzing contrastive stress in a multimodal context—specifically as *prosodic pointing*—for the teaching and learning of L2 prosodic pragmatics and the development of interpretive abilities in the L2 learner's mind. Her account sees contrastive stress as a tool which provides an extra cue to relevance theoretic stimulus ostension by altering the salience of one particular constituent in an utterance. In *Nutritional labelling, communication design, and relevance*, Scott adopts relevance theory notions as a means of explaining the relative effectiveness of three different nutrition labeling systems in communicating information and influencing consumer food choices. The relative success or failure of these labeling systems, Scott claims, are best explained in terms of the processing effort and inferential steps required from the consumer when accessing relevant contextual assumptions and deriving relevant implications in decision-making contexts. In other words, the success or failure of the various labeling systems is linked to their relevance in the context of interpretation.

## Author contributions

TW: Writing—original draft. KS: Writing—review & editing. DM: Writing—review & editing. CJ: Writing—review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## 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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fpsyg.2022.1073213



OPEN ACCESS

EDITED BY
Kate Scott,
Kingston University, United Kingdom

REVIEWED BY
Steve Oswald,
Université de Fribourg, Switzerland
Maria Grazia Rossi,
Universidade NOVA de Lisboa, Portugal

*CORRESPONDENCE
Francesca Bonalumi
✉ bonalumif@ceu.edu

†These authors have contributed equally to this work and share first authorship

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

RECEIVED 18 October 2022
ACCEPTED 05 December 2022
PUBLISHED 04 January 2023

CITATION
Bonalumi F, Bumin FB, Scott-Phillips T and Heintz C (2023) Communication and deniability: Moral and epistemic reactions to denials.
*Front. Psychol.* 13:1073213.
doi: 10.3389/fpsyg.2022.1073213



# Communication and deniability: Moral and epistemic reactions to denials

Francesca Bonalumi[1*†], Feride Belma Bumin[2†], Thom Scott-Phillips[3] and Christophe Heintz[1]

[1]Department of Cognitive Science, Central European University, Vienna, Austria, [2]Department of Economics and Business, Central European University, Vienna, Austria, [3]Institute for Logic Cognition Language and Information, University of the Basque Country, San Sebastian, Spain

People often deny having meant what the audience understood. Such denials occur in both interpersonal and institutional contexts, such as in political discourse, the interpretation of laws and the perception of lies. In practice, denials have a wide range of possible effects on the audience, such as conversational repair, reinterpretation of the original utterance, moral judgements about the speaker, and rejection of the denial. When are these different reactions triggered? What factors make denials credible? There are surprisingly few experimental studies directly targeting such questions. Here, we present two pre-registered experiments focusing on (i) the speaker's incentives to mislead their audience, and (ii) the impact of speaker denials on audiences' moral and epistemic assessments of what has been said. We find that the extent to which speakers are judged responsible for the audience's interpretations is modulated by their (the speakers') incentives to mislead, but not by denials themselves. We also find that people are more willing than we expected to revise their interpretation of the speaker's utterance when they learn that the ascribed meaning is false, regardless of whether the speaker is known to have had incentives to deceive their audience. In general, these findings are consistent with the idea that communicators are held responsible for the cognitive effects they trigger in their audience; rather than being responsible for, more narrowly, only the effects of what was "literally" said. In light of our findings, we present a new, cognitive analysis of how audiences react to denials, drawing in particular on the Relevance Theory approach to communication. We distinguish in particular: (a) the spontaneous and intuitive re-interpretation of the original utterance in light of a denial; (b) the attribution of responsibility to the speaker for the cognitive effects of what

is communicated; and (c) the reflective attribution of a particular intention to the speaker, which include argumentative considerations, higher-order deniability, and reputational concerns. Existing experimental work, including our own, aims mostly at (a) and (b), and does not adequately control for (c). Deeper understanding of what can be credibly denied will be hindered unless and until this methodological problem is resolved.

KEYWORDS

**deniability, Relevance Theory, strategic speaker, indirect communication, pragmatics, accountability**

## 1 Introduction

There is always a gap, however small, between what is linguistically encoded in a sentence and what is communicated by the speaker when using that sentence in context (Carston, 2004)—as Grice (1989) famously put, between "what is said" and "what is meant". As a consequence, it is always possible in theory to deny that what an audience has inferred was indeed what was actually meant. This in turn raises the prospect that deniability could be used in a strategic way, such that speakers generate indirect formulations when there is a risk that their intended meaning may cause an undesired response (Brown and Levinson, 1987; Pinker, 2007; Pinker et al., 2008; Lee and Pinker, 2010). Classic examples include acts of bribery ("I'd do that for anybody who needs a proper guidance."), sexual innuendos ("It's going to be a long night. [...] And I don't particularly like the book I've started.") and insinuations ("Handsome armour. Not a scratch on it.").[1] In practice, of course, some denials of intended meaning are far more credible than others. These possibilities raise important questions about commitment and credibility in language use. What factors make denials credible? What kind of cognitive reactions do audiences have to different types of denial? Such questions are important not only because answering them would shed light on the cognitive processes involved in communication; but also because denials and strategic use of language have a pervasive role in our daily life, and its consequent influence in domains such as politics and the law.

There are few experimental studies that directly target these issues, despite the important role that denials and deniability play in human interaction. Most of previous research on deniability and implicit communication consists of theoretical contributions, stemming from a widely spread assumption that denying what was implied should be more credible than denying what was explicitly expressed (Fricker, 2012; Camp, 2018). Deniability has been also taken to jeopardise the necessary public responsibilities ensuring that testimony provides reasons to believe the transmitted knowledge (Peet, 2015; Davies, 2019). Furthermore, deniability has been examined as a type of defence strategy that a speaker can appeal to in order to deny their commitment to an ascribed meaning (i.e., meaning initially ascribed by the audience) (Boogaart et al., 2020; see also Morency et al., 2008). Looking at the experimental literature, instead, there is, to our knowledge, little research targeting these research questions (for exceptions see Sternau et al., 2015; Reins and Wiegmann, 2021; Bonalumi et al., 2022). The most developed line of experimental research in this area targets not audience interpretations, but rather the specific situational conditions when it may be advantageous for speakers to exploit the possibility of denial for their own strategic ends (Lee and Pinker, 2010). By contrast, we know relatively little about the audience side. What cognitive reactions are triggered when a communicator denies having intended the meaning that the audience appears to have actually inferred? When are they triggered, and why?

Given the relative dearth of studies directly targeting audience reactions to denials, here we present two pre-registered experiments targeting the question of what cognitive effects denials can have (§2 and §3). We focus in particular on (i) the role of speaker's incentives to mislead their audience, and (ii) the impact of speaker's denials on their audiences' moral and epistemic reactions. We adopted this particular focus because these are cases where speakers' incentives to mislead their audience may be relevant factors in determining what audiences might infer. In particular, we reasoned that prior incentives to mislead the audience can provide background information that shapes the interpretation of a denial (Mazzarella, 2021), and thus impacts its credibility. Imagine that you have asked your daughter if she had finished her homework, and she confirms it. After checking what she had to do, you realise that she has not done the math homework due on Friday. When confronting her, she replies "I meant the homework for tomorrow!." How plausible is your daughter's denial? We hypothesised that her denial would be deemed less plausible if she had incentives to lie

1 The examples are respectively taken from scenes from the movies The Wolf of Wall Street (Scorsese, 2013), in which the broker Belford suggests that he could involve the FBI agent into a 500mln trade; North by Northwest (Hitchcock, 1959), in which Eve invites Roger to sleep together in her train couchette; and Game of Thrones (Benioff and Weiss, 2011), in which Ned Stark insinuates that Jamie Lannister is a coward.

or mislead you.[2] For instance, she might have been at the same time asking for the permission to go out with her friends. We hypothesised that if the audience is aware of these incentives, this will impact on the credibility of the denial, as measured by the interpretation and re-interpretation of utterances, and moral assessments of the speaker's communicative behaviour. Our results are partially consistent and partially inconsistent with these predictions. We find that the ascription of the speaker's responsibility is indeed modulated by their incentives to mislead their audience (§2). However, their denial did not make a difference. We also find that people are more willing than we expected to revise their interpretation of the speaker's utterance when they learn that the ascribed meaning is false, regardless of whether the speaker is known to have had incentives to deceive their audience (§3). This pattern of results is consistent with the idea that communicators are held responsible for the cognitive effects they trigger in their audiences.

In light of our results, we distinguish three types of cognitive processes that impact on how people react to a denial (§4). These are: (a) the spontaneous and intuitive re-interpretation of the original utterance in light of a denial; (b) the attribution of responsibility to the speaker for the cognitive effects of what is communicated; and (c) the reflective attribution of a particular communicative intention to the speaker, that is based on the evidence that one has to claim that that the speaker lied or intentionally misled their audience.

# 2 Experiment 1: Moral reaction

Experiment 1 was designed to test the hypothesis that the audience holds the speaker responsible for the cognitive effects of their communicative act in view of their (the speaker's) incentives to mislead the audience. When such incentives are present, the speaker's denial should be implausible, and as such it should not mitigate their (the speaker's) responsibility. Here, we operationalised responsibility as the social consequences that the speaker is called to pay in terms of moral blameworthiness.

We reasoned that a mitigation of the speaker's ascribed responsibility can be considered a reliable proxy for the audience's acceptance of the denial, and in turn, for the presence of plausible deniability. Thus, we measured the speaker's ascribed responsibility by asking the participants to rate the speaker's blameworthiness for misleading the audience. If blame ratings were negatively affected or not affected by the presence of a denial, i.e., if participants would maintain or increase the severity of their blameworthiness judgement in the presence of a denial, that would suggest that such denial was deemed not plausible.

We thus predicted that the speaker's incentives to mislead the audience would cause an increase in their perceived blameworthiness. We also predicted that the speaker's denial of the meaning the audience had initially ascribed to the utterance (hereafter, "ascribed meaning") would lead participants to blame the speaker with less severity if the speaker had no incentive to mislead the audience in the first place.

## 2.1 Methods

The study was pre-registered on Open Science Framework, with sample size, planned analyses and participants exclusion criteria specified. The pre-registration document is available at https://osf.io/jkn57.

### 2.1.1 Participants

A power analysis that was conducted with RStudio 1.1.463 (R Core Team, 2020) by using the "rsm" package in R (Harrell, 2022) showed that with 500 participants, assuming small to medium effect size, we would obtain approximately 92% of power when $\alpha = 0.05$. We thus planned to recruit 500 participants *via* Amazon MTurk (Amazon Mechanical Turk[3]), and 11 additional participants were also included in the analysis since they completed the survey before we closed the survey collector. Being above the age of 18 was the only criteria for participant selection. Each participant provided informed consent before the experiment and were paid $0.40 for their participation. Participants who failed the attention check were excluded ($N = 4$), thus the final sample resulted in 507 participants (236 females, 1 other, 1 prefer not to say, M_age = 39): 252 participants were assigned to the incentive condition (125 in the denial condition, 127 in the no-denial condition) and 255 participants were assigned to the no-incentive condition (128 in the denial condition, 127 in the no-denial condition).

The methods used in this and in the following study are in accordance with the international ethical requirements of psychological research and approved by the EPKEB (United Ethical Review Committee for Research in Psychology) in Hungary.

### 2.1.2 Materials

We created four different scenarios which followed the following structure:

- *Context part* depicted a social situation between a speaker and a listener, and included information about the speaker's

2 There is a sizeable literature on whether conveying indirectly false content should qualify as lying rather than merely misleading, both theoretical (e.g., Carson, 2006; Meibauer, 2014, 2018; Viebahn, 2017; Saul, 2018; Marsili, 2021), and experimental (e.g., Danziger, 2010; Wiegmann et al., 2016, 2021; Willemsen and Wiegmann, 2017; Antomo et al., 2018; Reins and Wiegmann, 2021; Viebahn et al., 2021). For this study we do not take any particular theoretical stance on what counts as lie; although people may entertain the belief that lying is morally worse than misleading (see Chisholm and Feehan, 1977; Adler, 1997), if only for argumentative reasons (see § 4).

3 https://www.mturk.com

TABLE 1 Example of scenarios structure and measures under four different conditions presented in Experiment 1.

| Context/Incentive | Context/No-incentive |
|---|---|
| Tommy and Thelma have been in a relationship for a few years. | Tommy and Thelma are siblings and have a very close relationship. |
| They live in the same college dorm. At the beginning of the new term, they meet a new student, Sara, in the dorm cafeteria. Tommy and Sara start spending a lot of time together, and Tommy knows that Thelma also likes Sara and is happy about them hanging out. One day Thelma looks for Tommy and cannot find him anywhere. | |
| **Dialogue** | |
| Thelma asks Tommy when he is back: "*Where were you? I couldn't find you anywhere.*" | |
| Tommy answers: "*Sorry, I went to the laundry room.*" | |
| **Negation** | |
| Later, some friends tell Thelma that Tommy and Sara were together that afternoon. | |
| **Denial** | **No-denial** |
| Thelma says to Tommy: "*I thought you said that you were doing your laundry.*" | **x** |
| Tommy answers: "*Oh no, I didn't say that. I just meant that I was helping Sara because she didn't know how to use the washing machine and asked me for help.*" | |
| **Attention check** | |
| Who was Thelma looking for that afternoon? | |
| • Tommy/Sara/Nobody | |
| **Blame question** | |
| If you were Thelma, how much would you blame Tommy for misleading you? | |
| • 1 (Not at all)/2/3/4/5/6 (Completely) | |

incentives to mislead the audience: in the incentive condition, the speaker was described as having incentives to let their listener believe something, whereas in the no-incentive condition the speaker did not have any incentive or have disincentives to do so.

- *Dialogue part* included the question of the listener about an event; the speaker's response to the question was an utterance (X) which yielded an ascribed meaning (Y). The dialogue part was identical in all conditions.
- *Negation part* included the information that the listener realises that Y is false. This part was identical in all conditions.
- *Denial part* included the dialogue in which the listener confronts the speaker by stating that "*I thought you said Y*" and the speaker denied having meant that Y (ascribed meaning), and offered an alternative intended meaning (Z) for their utterance X: "*I didn't say that. I just meant Z*". Then, the denial part was present only in the denial condition.

As suggested by Mazzarella (2021), we constructed the speakers' denials as including the offer of an alternative interpretation for the utterance X. **Table 1** shows one specific example. All stimuli are available at https://osf.io/bmqk4/.

### 2.1.3 Procedure and design

Experiment 1 used a 2 × 2 between-subject design.[4] The factors were "incentives" (incentive vs. no-incentive) and "denial" (denial vs. no-denial). Participants were randomly presented with one unique scenario manipulated according to one of four different conditions: incentive and denial, incentive and no-denial, no-incentive and denial, and no-incentive and no-denial.

After reading scenarios, participants responded to two questions: an attention check, which was a multiple-choice question designed to check the reliability of the participant's answer, and a blame question, which was a 6-point Likert scale question designed to measure the moral reaction of the participant (see **Table 1**). The attention check was different for each scenario regarding the context while the blame question was the same for every scenario under all conditions; "*If you were listener, how much would you blame the speaker for misleading you?*" [1: Not at all, 2, 3, 4, 5, 6: Completely].

We expected that both the "incentives" and the "denial" factors, as well as their interaction, would cause a significant effect on the moral reaction of the participants. We predicted that participants would blame the speaker in the incentive condition more than in the no-incentive condition, and they would blame the speaker less in the denial condition than in the no-denial condition. However, if the denial was not deemed plausible, as we reasoned would be the case in the incentive condition, we predicted that in the denial condition participants would blame the speaker more, or at least not less, than in the no-denial condition.

## 2.2 Data analysis

In our pre-registered analyses, we planned to use an ordered logistic regression model to test our hypothesis and to include

4 We opted for a between-subject design because we reasoned that a within-subject design could trigger an experimenter effect if participants were offered both conditions of the same scenario.

TABLE 2 Results of the multilevel ordered logistic regression model.

| Variable | β | SE (β) | *p*-value | Odds ratio | 95% CI of odds ratio | |
|---|---|---|---|---|---|---|
| | | | | | Lower bound | Upper bound |
| Incentives | | | | | | |
| **Yes**** | **0.620**** | **0.229** | **0.007** | **1.858** | **1.185** | **2.912** |
| No | Reference | | | | | |
| Denial | | | | | | |
| Yes | 0.221 | 0.223 | 0.321 | 1.248 | 0.806 | 1.931 |
| No | Reference | | | | | |
| Incentives x Denial | | | | | | |
| Yes | −0.509 | 0.320 | 0.112 | 0.601 | 0.302 | 1.126 |
| No | Reference | | | | | |

$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
Bold values indicate significant variables with $p$-value < 0.05.

the "scenario" variable as a random factor in case the descriptive statistics showed different distributions of responses across scenarios.[5] Before the analysis, we thus checked the distributions of participants' responses and detected such difference across scenarios.[6] Thus, we added "scenario" as a random factor and switched to multilevel ordered logistic regression model that is a significantly better fit compared to a model with the intercept only, $\chi^2$ (4, $N$ = 507) = 145.64, $p < 0.001$.

## 2.3 Results

We ran our multilevel ordered logistic regression in Stata 17 (StataCorp, 2021). The results in **Table 2** show that speakers' blameworthiness was modulated by their prior incentives to mislead, but surprisingly not by the denials themselves nor by the interaction between the speakers' incentives and denials.

The estimated odds ratio of prior incentives points out that participants tended to blame the speakers 1.858 times more (95% CI: 1.185–2.912) by rating higher when the speakers had prior incentives to mislead the audience. However, we could not observe any significant effect of denial on the participants' rating level.

Consistent with our prediction, our results reveal that participants' moral judgements were sensitive to the speaker's incentives: participants blamed the speakers for a false ascribed meaning significantly more severely when speakers had incentives to do so (see **Figure 1**). On the other hand, contrary to our prediction, the effects of denial and its interaction with speaker's incentives were statistically insignificant. The fact that speakers denied the ascribed meaning did not affect the participants' judgments, regardless of whether the speaker denied an ascribed meaning that they had or had not an incentive to convey in the first place. These findings indicate that speakers, with incentives to mislead their audience, paid higher social costs and were held more blameworthy for misleading their audience, regardless of whether they denied having meant the falsely ascribed meaning.

We assumed that ascribing responsibility to the speaker for having misled the audience, i.e., judging them as blameworthy, is mediated by an interpretation process about speaker's meaning; and hence that when a denial is offered, the ascription of responsibility should be mediated by a re-interpretation process about speaker's meaning (as suggested by Mazzarella, 2021). To confirm this, we conducted Experiment 2.

# 3 Experiment 2: Epistemic reaction

A speaker's denial has an impact on other cognitive processes beyond the ascription of responsibility (see also §4). If an ascribed meaning is plausibly deniable, its denial should first lead to re-interpretation. We designed Experiment 2 to test the hypothesis that the speaker's prior incentives to mislead the audience are salient contextual assumptions that might block the audience's re-interpretation of the speaker's intended meaning.

We measure perceived speaker meaning by asking participants an *interpretation* question. The interpretation question offers two possible interpretations for what the speaker meant (the initially ascribed meaning and the alternative meaning offered with the denial), with degrees of uncertainty. If the participants choose the ascribed meaning over the alternative meaning more frequently when the speaker's incentives are present than when they are not, this would

5 see https://osf.io/bmqk4/

6 Our model fits the data better compared to the same model without random effect, $\chi^2$ (1, $N$ = 507) = 140.04, $p < 0.001$, and improves the significance level and estimated odds ratio of our already significant variables as results show in **Table 2**. See https://osf.io/bmqk4/for more details.

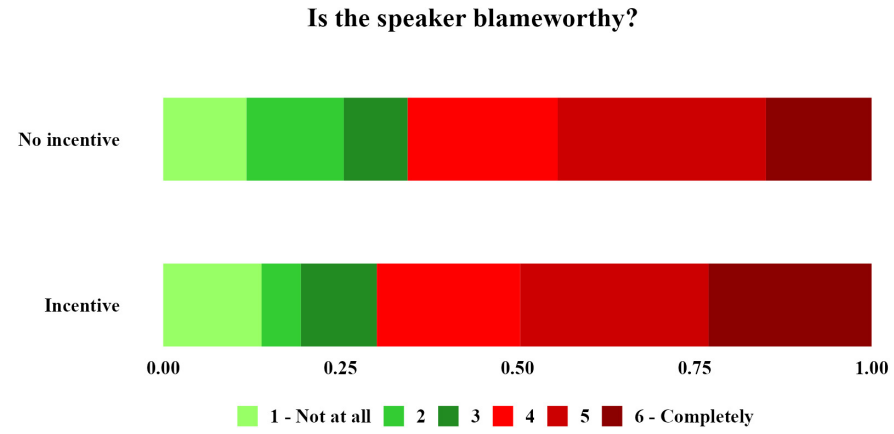



FIGURE 1
Distribution of participants' total responses to the blame question. There is a shift towards greater blame in the incentive condition. That is, participants tended to blame the speaker significantly more for misleading the audience when the speaker had prior incentives to do so, independently of the presence of their denial.

suggest that the speaker's prior incentives affect the re-interpretation of what the speaker meant. Additionally, if the participants choose the alternative meaning more often when denial is present, but not when it is absent, that would suggest the denial was plausible. We hence predicted that when the speaker had prior incentives to mislead the audience their denial would not be deemed as plausible, and thus it would not lead to an update of the interpretation of the speaker's utterance; whereas when the speaker did not have any incentives to mislead their audience, their denial would lead to a re-interpretation of the speaker's utterance in favour of the alternative proposed.

## 3.1 Methods

The study was pre-registered on OSF.io, with sample size, planned analyses and participants exclusion criteria specified. The pre-registration document is available at https://osf.io/jkn57.

### 3.1.1 Participants

A statistical power of 95% with $\alpha = 0.05$ was calculated by the "Basic Functions for Power Analysis (pwr)" package (Champely, 2020) of RStudio 1.4.1103 (R Core Team, 2020) for 752 participants with a small effect size. 790 participants attended the experiment before we closed the experiment and they were recruited *via* Amazon MTurk (Amazon Mechanical Turk[7]). The only criteria for participants was being above 18. Each participant was asked for their consent before starting the experiment and was compensated with $0.40. We excluded data from participants who failed the attention check question ($N = 38$), resulting in 752 participants (360 females, 1 others, 1 prefer not to say, M_age = 40.20); 370 in the incentive condition (121 in the negation and denial condition, 123 in the negation and no-denial condition, 126 in the no-negation condition) and 382 in the no incentive condition (128 in the negation and denial condition, 123 in the negation and no-denial condition, 131 in the no-negation condition).

### 3.1.2 Materials

To maintain comparability with the previous study, we used the same four scenarios that were used in Experiment 1. However, contrary to Experiment 1, we additionally manipulated the negation part in order to isolate the effect of denial on re-interpretation. The scenarios followed the structure below.

- *Context part* depicted a social situation between a speaker and a listener, and included information about the speaker's incentives to mislead the listener.
- *Dialogue part* included a question of the listener about an event; the speaker's response to the question was an utterance (X) which yielded an ascribed meaning (Y). The dialogue part was identical in all conditions.
- *Negation part* included the information about the listener realising that Y is false. The evidence that Y is false was present in the negation condition and absent in the no-negation condition.
- *Denial part* included the dialogue in which the listener confronts the speaker by stating that "I thought you said Y" and the speaker denied having meant that Y (ascribed meaning), and offered an alternative meaning (Z) for their utterance X: "*I didn't say that. I just meant Z*". As in Experiment 1, the denial part was present only in the denial condition.

**Table 3** shows an example. Again, all stimuli are available at https://osf.io/bmqk4/.

TABLE 3 Example of the scenario structure and measures used in the Experiment 2.

| Context–incentive | Context–no-incentive |
|---|---|
| Tommy and Thelma have been in a relationship for a few years. | Tommy and Thelma are siblings and have a very close relationship. |
| They live in the same college dorm. At the beginning of the new term, they meet a new student, Sara, in the dorm cafeteria. Tommy and Sara start spending a lot of time together, and Tommy knows that Thelma also likes Sara and is happy about them hanging out. One day Thelma looks for Tommy and cannot find him anywhere. | |
| **Dialogue** | |
| Thelma asks Tommy when he is back: *"Where were you? I couldn't find you anywhere."* | |
| Tommy answers: *"Sorry, I went to the laundry room."* | |
| **Negation** | **No-negation** |
| Later, some friends tell Thelma that Tommy and Sara were together that afternoon. | **x** |
| **Denial** | **No-denial** |
| Thelma says to Tommy: *"I thought you said that you were doing your laundry."* | **X** |
| Tommy answers: *"Oh no, I didn't say that. I just meant that I was helping Sara because she didn't know how to use the washing machine and asked me for help."* | |
| **Attention check** | |
| Who was Thelma looking for that afternoon? | |
| • Tommy/Sara/Nobody | |
| **Interpretation question** | |
| When Tommy said "I went to the laundry room", did Tommy mean he was doing his laundry or he was helping Sara? | |
| ❖ Tommy clearly meant he was doing his laundry | |
| ❖ Tommy probably meant he was doing his laundry | |
| ❖ what Tommy meant is unclear | |
| ❖ Tommy probably meant he was helping Sara | |
| ❖ Tommy clearly meant he was helping Sara | |

Bold values indicate significant variables with $p$-value < 0.05.

### 3.1.3 Procedure and design

Experiment 2 used a 3 × 2 between-subjects design. The factors were "incentives" (incentive vs. no incentive), "denial" (denial and no-denial), and "negation" (negation and no-negation). Participants were randomly presented with one unique scenario manipulated according to one of six different conditions: incentive and denial, no-incentive and denial, incentive and negation, no-incentive and negation, incentive and no-negation, no-incentive and no-negation.

After reading the scenario, participants responded to two questions: the attention check, and the interpretation question, which was a multiple-choice question with five independent levels and designed to check which meaning is understood to be conveyed by the speaker, the ascribed or the alternative meaning, and with how much certainty. The interpretation question was the same for every scenario under all conditions: "*When [speaker] said utterance (X), did [speaker] mean ascribed meaning (Y) or alternative meaning (Z)?*" [the speaker clearly meant the ascribed meaning (Y); the speaker probably meant the ascribed meaning (Y); what the speaker meant is unclear; the speaker probably meant the alternative meaning (Z); the speaker clearly meant the alternative meaning (Z)].

We hypothesised that the speaker's incentives to mislead the audience would impact the plausibility of their denial, thus we expected an interaction between the "denial" and the "incentives" factors. We further expected that the "negation" factor may have an additional significant effect alone on participants' responses without the presence of the speaker's denial. We thus predicted:

- A significant effect of "incentives": participants would choose the alternative meaning more often in the no-incentive conditions than in the incentive conditions.
- An interaction between "denial" and "incentives" factors: participants would choose the alternative meaning more often in the denial condition than in the no denial condition, but only in the no-incentive conditions.
- A significant effect of "negation": participants would choose the alternative meaning more often in the negation conditions than in the no-negation condition.

## 3.2 Data analysis

To test our hypothesis, we pre-registered that we will use a multinomial logistic regression model to analyse a categorical dependent variable, i.e., participants' responses to interpretation question in our model, with more independent factors, i.e., "incentives", "negation" and "denial" factors.[7] We

7 In pre-registration, we mentioned a gender effect that was observed while piloting the study. When, we added the "gender" variable to both of

ran two separate models with the same dependent but different independent variables for the ease of the analysis: (1) a denial model that included the "incentives" factor, the "denial" factor, and their interaction, (2) a negation model that included the 'incentives' factor, the 'negation' factor, and their interaction. In both of our models, we chose the "speaker clearly meant the ascribed meaning" level of dependent variable as our base category value. Both of our models improved their fit when we added the 'scenario' as a random effect and switched to multilevel multinomial logistic regression, denial model, $\chi^2$ (1, $N$ = 506) = 5.28, $p$ = 0.022, and negation model, $\chi^2$ (1, $N$ = 503) = 13.30, $p$ = 0.001. Additionally, both of our models fit significantly better compared to model with the intercept only; denial model, $\chi^2$ (13, $N$ = 506) = 92.74, $p < 0.001$, and negation model, $\chi^2$ (13, $N$ = 503) = 79.03, $p < 0.001$.

## 3.3 Results

We ran both of our multilevel multinomial logistic regression models in Stata 17 (StataCorp, 2021). The results of both models are shown in **Table 4**.

Contrary to our prediction, the speaker's incentives to mislead the audience did not affect participants' interpretation of the intended meaning overall. No significant effect of "incentives" was found. However, the presence of denial did have a significant effect on the participants' responses. This suggests that participants are disposed to think that they misinterpreted the intended meaning and to accept the alternative as the originally intended meaning. Thus, the denial model did not confirm our prediction regarding the effect of the speaker's incentives on their re-interpretation process. Our results suggest that, as proposed by Mazzarella (2021), the presence of a speaker's denial triggers a re-interpretation process.

Also, as we predicted, when participants were provided with the information that the ascribed meaning was false, this affected participants' assessments of the intended meaning (see **Figure 2**). The new information caused participants to update their belief about what the speaker intended to communicate, even when they were not provided with the speaker's denial of the ascribed meaning.

Collectively, these results show that people are able and willing to retrospectively ascribe a different informative intention to the speaker. When presented with relevant information such as a denial or a negation of their ascribed meaning, the re-interpretation process occurs. Perhaps surprisingly, we did not find evidence that the interpretation and re-interpretation processes are sensitive to the speaker's incentives to mislead the audience; participants were rather inclined to revise their interpretation in both situations.

our models, we did not detect any significant gender effect (in the denial model, $\chi^2$ (4, $N$ = 506) = 3.49, $p$ = 0.480, 2); in the negation model, $\chi^2$ (4, $N$ = 503) = 0.16, $p$ = 0.997). Also, we did not have any pre-assumption on why gender should have an effect, so we excluded 'gender' from our models.

# 4 "That's not what i meant!" rethinking deniability

Our findings show that a satisfactory account of plausible deniability relies on disentangling multiple facets of audiences' reactions to denials. In particular, the dissociation we observed between moral (§2) and epistemic (§3) reactions towards denials was unpredicted and is puzzling. While participants' moral reaction have been found to be influenced by the speaker's incentives to mislead the audience, but not by their (the speaker's) denial, the opposite was found for participants' epistemic reactions; that is, participants' reported willingness to re-interpret speaker meaning was influenced by their (the speaker's) denials, and other evidence that the initially ascribed meaning was false, but not by their incentives to mislead. In light of these findings, here we re-analyse deniability, making distinctions that have not been clearly made in the previous literature on the topic. These distinctions are inspired in particular by the Relevance Theory approach to communication and cognition (e.g., Wilson and Sperber, 2012; Mazzarella, 2021; Heintz and Scott-Phillips, 2022), but could also be derived from other theoretical frameworks.

It is essential to distinguish a communicator's intended meaning from what could be called the "ascribed" meaning; that is, the meaning the audience ascribes to the utterance. Denials are statements from the communicator about how the meaning ascribed by the audience differs from the communicator's intended meaning. Such mismatches between intended meaning and ascribed meaning occur all the time in ordinary communication, and humans have developed and use a wide array of mechanisms for "repairing" dialogue when this occurs (Dingemanse and Enfield, 2015; Dingemanse et al., 2015). These include interjections such as "Huh?" and "What?"; question words seeking clarification; partial repeats of the source of uncertainty followed by a question word; reformulations of what was meant; and others. However, some of the time, denials trigger further cognitive reactions in audiences that go beyond repair, and corresponding clarification of what the communicator had originally meant.

At least three possible reactions should be distinguished. These are not mutually exclusive, and will in some cases co-occur with one another.

> *a. Audiences may re-interpret the original utterance, potentially in line with the new interpretation offered by the speaker.*

Denials often are accompanied by an alternative interpretation aimed to trigger a re-interpretation process

TABLE 4 Effects of the "incentives", "denial" and "negation" factors on participants' responses to the interpretation question in the multilevel multinomial regression denial model and negation model.

| Model | Effect | Model fitting criteria | Likelihood ratio tests | | |
|---|---|---|---|---|---|
| | | −2 Log likelihood of reduced model | Chi-squared | Degrees of freedom | *p*-value |
| Denial model | Intercept | 1522.12 | 0.000 | 0 | – |
| | Incentives | 1519.95 | 2.17 | 4 | 0.705 |
| | **Denial*** ** | **1484.06** | **38.06** | **4** | **0.000** |
| | Incentives*Denial | 1521.24 | 0.88 | 4 | 0.927 |
| Negation model | Intercept | 1494.96 | 0.000 | 0 | – |
| | Incentives | 1492.79 | 2.17 | 4 | 0.704 |
| | **Negation*** ** | **1466.37** | **28.59** | **4** | **0.000** |
| | Incentives*Negation | 1491.65 | 3.31 | 4 | 0.507 |

$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
Bold values indicate significant variables with $p$-value < 0.05.

in the audience. For instance, in order to deny to have lied with his infamous statement "*I did not have sexual relations with that woman*", former US president Bill Clinton offered the alternative interpretation for which the definition of "sexual relations" was thought not to include the specific interactions he admittedly had with Lewinsky. Reinterpretations typically require increasing the saliency of some contextual assumptions that were neglected when the communicative act was initially produced. If the new alternative interpretation meets a better trade-off in terms of cognitive utility compared to the old interpretation, then the re-interpretation successfully occurs and the denial can be perceived as plausible (Mazzarella, 2021). In our Experiment 2, participants reported a willingness to re-interpret the communicator's original utterances in this way. The output of this process can be described as a type of intuitive belief, because it consists in inferences that are spontaneous, implemented by our communicative capacities. Specifically, the belief about re-interpretation is not based on an assessment of the reasons for forming such belief (on the difference between intuitive and reflective beliefs, see Sperber, 1997; Mercier and Sperber, 2019).

> *b. Audiences may ascribe responsibility to speakers for the cognitive effects caused by their earlier communicative act, especially when the audience had relied on those cognitive effects.*

In general, speakers are held accountable for the cognitive effects caused by their communicative acts, rather than being responsible for, more narrowly, only the effects of what was "literally" said (Morency et al., 2008; Haugh, 2013; Bonalumi et al., 2020; Yuan and Lyu, 2022). These accountability effects may be particularly sensitive to the plausibility of the denial. The fact that the speaker suggests that there is a mismatch between the intended meaning and the ascribed meaning may mitigate blameworthiness, but can also backfire if the denial is not plausible, and even more so when speakers deny having intended these cognitive effects (Bonalumi et al., 2022; see also Oswald, 2022, for a similar take on insinuations). Bill Clinton's denial attempt certainly was not convincing and triggered additional public outrage. When speakers attempt to eschew the responsibility for the cognitive effects they had generated in the audience, then the audience's moral evaluation of the speaker is impacted.

> *c. Audiences may reflexively accept or reject a denial, considering the evidence they possess that the speaker really meant the ascribed meaning. These reflective beliefs include anticipating an argument with the speaker about what he or she truly intended.*

The audience's confidence in explicitly or publicly attributing an informative intention to the speaker is informed not only by their intuitive interpretation of the utterance, but also by considerations of the evidence in favour of attributing to the speaker the ascribed meaning, i.e., the denied meaning, or the alternative meaning. The belief about the speaker's actual intended meaning is, in that case, a reflective belief. In fact, the audience may reflect that meanings that are "literally" expressed as less deniable than meanings that are implied: this is because the uttered words can easily be used as good evidence for ascribing literal meanings. By contrast, implied meanings are (intuitively) attributed on numerous contextual cues that might be harder to use as evidence in an argument about intended meaning. More generally, reasoning on ascribed meaning is likely to ensure that denials are accepted more often than not because of the widely shared intuitions that one has a privileged access to one own's thoughts and intentions, and thus it is not possible to really know what others think (Keane, 2008; Astuti, 2015; see also Burge and Peacocke, 1996)—or that language is a digital medium that encodes meaning (Pinker, 2004). The reflexive acceptation or rejection of denial thus involve other considerations such as higher-order deniability,

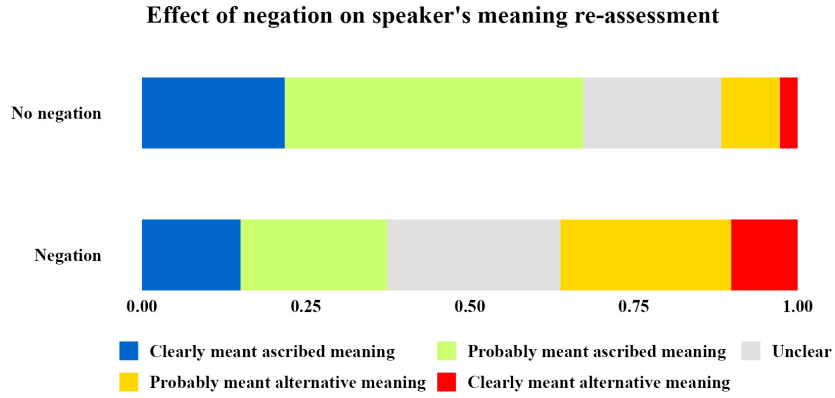



FIGURE 2
Distribution of participants' total responses to the interpretation question, as the measure of effect of the "negation" factor. As we can see, when provided with the information about the ascribed meaning being false, participants were significantly more likely to re-interpret speaker meaning.

possible deniability, and relationship management concerns (Lee and Pinker, 2010; Dinges and Zakkou, in press; see also Elder, 2021, for a discussion on deniability and micro-aggression). Additionally, the audience's confidence in the ascribed meaning can be informed by further contingent elements that suggest that engaging in an overt reproach of the speaker will or will not be successful (e.g., power relations, appropriateness of the reproach, etc.) (see also Dinges and Zakkou, in press). In particular, both speaker's reputation and their institutional roles are factors that we did not manipulate here, but we expect them to affect significantly the perceived plausibility of a denial and as such they should be explored in future research.

These three possible cognitive reactions can be present simultaneously. Consider, for example, the famous yacht scene from "The Wolf of Wall Street"[8]: Jordan Belford (Leonardo DiCaprio) implicitly attempts to bribe an FBI agent (Kyle Chandler) by stating "*I'd do that [providing information about a millionaire stock trade] for anybody who needs a proper guidance.*" Once confronted by the FBI agent, ("*You just tried to bribe a federal officer.*"), Belford denies having had such intention ("*I don't know what you're talking about.*"). The FBI agent appears to hold the intuitive belief that Jordan Belford's denial is implausible, maintaining the inferred informative intention of a bribe proposal ("*C'mon you know what I'm talking about.*"), but at the same time may lack the confidence to publicly attributing this informative intention following Jordan's denial, in particular in front of third parties—and indeed the immediate accusation of bribe is not acted upon (as Belford points out, "*That would not hold up in a court of law.*"). During the whole interaction, in any case, the FBI agent is clearly taking a moral stance against the broker (e.g., "*You, Jordan, you got this way all on your own—Good for you little man.*"), which is of course unaffected by his (Belford)'s denial attempt.

Existing experimental work, including our own, does not adequately control for judgements informed by reflective beliefs about possible deniability, i.e., possibility (c) above. In the two studies we have presented here, we investigated the effect of speaker's denials and speaker's incentives to mislead the audience on moral and epistemic evaluations. As we reported in §2–3, our results are only partially in line with these predictions. We suggest this may be because our experimental designs have conflated reaction (c) with either of the other two. Our intention was to target reactions (a), intuitive reinterpretation, and (b), responsibility ascription. However, we may have additionally triggered reaction (c), reflective beliefs, in particular argumentative considerations and participants' judgements about possible deniability instead.

Partly because we expected that epistemic and moral reactions would be consistent, we reason that the mismatch that we found between the reported re-interpretations and the ascriptions of responsibility may be due to the interference of such reflective beliefs. More specifically, we suggest that participants' reported epistemic judgements (a) could have been conflated with other argumentative considerations (c). Such argumentative considerations would have prevented participants from engaging in an explicit accusation about speaker's intentions. However, these considerations may have had less impact on participants' moral judgements (b); or they may even have been consistent with such judgements. In fact, and regardless of their actual intention to mislead, the favourable outcome for the speaker (i.e., their incentives) is good enough evidence for defending an explicit disapproval of the speaker's incompetent behaviour.

The difficulty of keeping these different reactions apart in experimental design is one that may have recurred in other

8 See https://www.youtube.com/watch?vojWbip26nQs

recent research on deniability (Sternau et al., 2015, 2017; Bonalumi et al., 2022). Future experimental designs must focus on operationalising plausible deniability in a way that tears apart (c) from (a) and (b). A deeper understanding of the diverse range of ways in which people react to denials will be hindered unless and until this methodological problem is resolved.

The important general point is that plausible deniability involves strategic cognition for both speakers and audiences. The speaker attempts not to produce evidence to be accused of lying, while the audience assesses whether the speaker has or had the intention to mislead. Thus, while audience may modulate their (re-)interpretation of what is said in view of the speaker's intentions, discussing a denial involves not only re-interpretation of the speaker's informative intention, but engagement in discussion or argument about those intentions.; and since speakers can always claim privileged access to their own past intentions, the audience may strategically avoid this outcome. A good cognitive description of plausible deniability must account for these different processes.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/jkn57.

## Ethics statement

The studies involving human participants were reviewed and approved by United Ethical Review Committee for Research in Psychology (EPKEB). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

FB: conceptualization, methodology, validation, data curation, writing – original draft, writing – review and editing, visualization, and project administration. FBB: software, validation, formal analysis, investigation, data curation, writing – original draft, visualization, and funding acquisition. TS-P: validation, writing – original draft, and writing – review and editing. CH: conceptualization, resources, writing – review and editing, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

This publication was the result of research conducted for Central European University, Private University. It was made possible by the CEU Open Access Fund. The research supporting this article was sponsored by Central European University Foundation of Budapest (CEUBPF), Grant serial Number: BPF/19188.

## Acknowledgments

We would like to thank our colleagues from Adaptive Cognition and Economics in Societies at CEU. We are further grateful to Nausicaa Pouscoulous and Diana Mazzarella for the inspiring conversations on the topic.

## 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.

## Publisher's note



## References

Adler, J. (1997). Lying, deceiving, or falsely implicating. *J. Philos.* 94, 435–452. doi: 10.2307/2564617

Antomo, M., Müller, S., Paul, K., Paluch, M., and Thalmann, M. (2018). When children aren't more logical than adults: An empirical investigation of lying by falsely implicating. *J. Pragmat.* 138, 135–148. doi: 10.1016/j.pragma.2018.09.010

Astuti, R. (2015). Implicit and explicit theory of mind. *Anthropol. Century* 13, 636–650.

Benioff, D., and Weiss, D. B. (Directors) (2011). *Game of thrones*. New York, NY: HBO.

Bonalumi, F., Mahr, J. B., Marie, P., and Pouscoulous, N. (2022). Beyond the implicit/explicit dichotomy: The pragmatics of commitment, accountability, and plausible deniability. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/z2bqt

Bonalumi, F., Scott-Phillips, T., Tacha, J., and Heintz, C. (2020). Commitment and Communication: Are we committed to what we mean, or what we say? *Lang. Cogn.* 2, 1–25. doi: 10.1017/langcog.2020.2

Boogaart, R., Jansen, H., and van Leeuwen, M. (2020). Those are your Words, Not Mine!" defence strategies for denying speaker commitment. *Argumentation* 35, 209–235. doi: 10.1007/s10503-020-09521-3

Brown, P., and Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Burge, T., and Peacocke, C. (1996). Our entitlement to self-knowledge: II. Christopher Peacocke: Entitlement, self-knowledge and conceptual redeployment. *Proc. Aristotel. Soc.* 96, 117–158. doi: 10.1093/aristotelian/96.1.117

Camp, E. (2018). "Insinuation, common ground, and the conversational record," in *New work on speech acts*, Vol. 1, (Cambridge: Oxford University Press). doi: 10.1093/oso/9780198738831.003.0002

Carson, T. L. (2006). The definition of lying. *Noûs* 40, 284–306. doi: 10.1111/j.0029-4624.2006.00610.x

Carston, R. (2004). "Truth-conditional content and conversational implicature," in *The semantics/pragmatics distinction*, ed. C. Bianchi (Stanford, CA: CSLI Publication).

Champely, S. (2020). *Basic functions for power analysis (1.3-0).* Available online at: https://github.com/heliosdrm/pwr (accessed May 1, 2021).

Chisholm, R. M., and Feehan, T. D. (1977). The intent to deceive. *J. Philos.* 74, 143–159. doi: 10.2307/2025605

Danziger, E. (2010). On trying and lying: Cultural configurations of Grice's maxim of quality. *Intercult. Pragmat.* 7, 199–219. doi: 10.1515/iprg.2010.010

Davies, A. (2019). Testimony, recovery, and plausible deniability: A response to Peet. *Episteme* 16, 18–38. doi: 10.1017/epi.2017.15

Dingemanse, M., and Enfield, N. J. (2015). Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguist.* 1, 96–118. doi: 10.2478/opli-2014-0007

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (2015). Universal principles in the repair of communication problems. *PLoS One* 10:e0136100. doi: 10.1371/journal.pone.0136100

Dinges, A., and Zakkou, J. (in press). On Deniability. *Mind*.

Elder, C.-H. (2021). Microaggression or misunderstanding? Implicatures, inferences and accountability. *J. Pragmat.* 179, 37–43. doi: 10.1016/j.pragma.2021.04.020

Fricker, E. (2012). Stating and insinuating. *Aristotel. Soc.* 86, 61–94. doi: 10.1111/j.1467-8349.2012.00208.x

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Harrell, F. E. J. (2022). *Regression modeling strategies (6.3-0).* Available online at: https://CRAN.R-project.org/package=rms

Haugh, M. (2013). Speaker meaning and accountability in interaction. *J. Pragmat.* 48, 41–56. doi: 10.1016/J.PRAGMA.2012.11.009

Heintz, C., and Scott-Phillips, T. (2022). Expression unleashed: The evolutionary & cognitive foundations of human communication. *Behav. Brain Sci.* 1–46. doi: 10.1017/S0140525X22000012

Hitchcock, A. (Director) (1959). *North by Northwest*. Beverly Hills: Metro-Goldwyn-Mayer.

Keane, W. (2008). Others, other minds, and others' theories of other minds: An afterword on the psychology and politics of opacity claims. *Anthropol. Q.* 81, 473–482. doi: 10.1353/anq.0.0000

Lee, J. J., and Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychol. Rev.* 117, 785–807. doi: 10.1037/a0019688

Marsili, N. (2021). Lying, speech acts, and commitment. *Synthese* 199, 3245–3269. doi: 10.1007/s11229-020-02933-4

Mazzarella, D. (2021). I didn't mean to suggest anything like that!": Deniability and context reconstruction. *Mind Lang.* 1–19. doi: 10.1111/mila.12377

Meibauer, J. (2014). *Lying at the semantics-pragmatics interface*. Berlin: De Gruyter Mouton. doi: 10.1515/9781614510840

Meibauer, J. (2018). The linguistics of lying. *Annu. Rev. Linguist.* 4, 357–377. doi: 10.1146/annurev-linguistics-011817-045634

Mercier, H., and Sperber, D. (2019). *The enigma of reason*. Cambridge, MA: Harvard University Press.

Morency, P., Oswald, S., and De Saussure, L. (2008). Explicitness, implicitness and commitment attribution: A cognitive pragmatic approach. *Belgian J. Linguist.* 22, 197–219. doi: 10.1075/bjl.22.10mor

Oswald, S. (2022). Insinuation is committing. *J. Pragmat.* 198, 158–170. doi: 10.1016/j.pragma.2022.07.006

Peet, A. (2015). Testimony, pragmatics, and plausible deniability. *Episteme* 12, 29–51. doi: 10.1017/epi.2014.31

Pinker, S. (2004). *The language instinct: How the mind creates language (Reprint edition)*. London: Penguin books.

Pinker, S. (2007). The evolutionary social psychology of off-record indirect speech acts. *Intercult. Pragmat.* 4, 437–461. doi: 10.1515/IP.2007.023

Pinker, S., Nowak, M. A., and Lee, J. J. (2008). The logic of indirect speech. *Proc. Natl. Acad. Sci. U.S.A.* 105, 833–838. doi: 10.1073/pnas.0707192105

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reins, L. M., and Wiegmann, A. (2021). Is lying bound to commitment? Empirically investigating deceptive presuppositions, implicatures, and actions. *Cogn. Sci.* 45:e12936. doi: 10.1111/cogs.12936

Saul, J. (2018). "Dogwhistles, political manipulation, and philosophy of language," in *New work on speech acts*, (Oxford: Oxford University Press). doi: 10.1093/oso/9780198738831.003.0013

Scorsese, M. (Director) (2013). *The wolf of wall street*. Los Angeles, CA: Paramount Pictures.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind Lang.* 12, 67–83. doi: 10.1111/j.1468-0017.1997.tb00062.x

StataCorp (2021). *Stata: Release 17 (Release 17)*. College Station, TX: StataCorp LLC.

Sternau, M., Ariel, M., and Giora, R. (2017). "Deniability and explicatures," in *Doing pragmatics interculturally*, eds R. Giora and M. Haugh (Berlin: De Gruyter Mouton), 97–120.

Sternau, M., Ariel, M., Giora, R., and Fein, O. (2015). Levels of interpretation: New tools for characterizing intended meanings. *J. Pragmat.* 84, 86–101. doi: 10.1016/j.pragma.2015.05.002

Viebahn, E. (2017). Non-literal lies. *Erkenntnis* 82, 1367–1380. doi: 10.1007/s10670-017-9880-8

Viebahn, E., Wiegmann, A., and Willemsen, P. (2021). Can a question be a lie? An empirical investigation. *Ergo* 8:7.

Wiegmann, A., Samland, J., and Waldmann, M. R. (2016). Lying despite telling the truth. *Cognition* 150, 37–42. doi: 10.1016/j.cognition.2016.01.017

Wiegmann, A., Willemsen, P., and Meibauer, J. (2021). Lying, Deceptive Implicatures, and Commitment. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/n96eb

Willemsen, P., and Wiegmann, A. (2017). "How the truth can make a great lie: An empirical investigation of the folk concept of lying by falsely implicating," in *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, (London), 3516–3621.

Wilson, D., and Sperber, D. (2012). *Meaning and relevance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107415324.004

Yuan, W., and Lyu, S. (2022). Speech act matters: Commitment to what's said or what's implicated differs in the case of assertion and promise. *J. Pragmat.* 191, 128–142. doi: 10.1016/j.pragma.2022.01.012

frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fcomm.2023.1088861



OPEN ACCESS

EDITED BY
Didier Maillat,
Université de Fribourg, Switzerland

REVIEWED BY
Caroline Jagoe,
Trinity College Dublin, Ireland
Johanna Miecznikowski,
University of Italian Switzerland, Switzerland

*CORRESPONDENCE
Magali A. Mari
✉ magali.mari@unine.ch

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

RECEIVED 03 November 2022
ACCEPTED 13 February 2023
PUBLISHED 03 March 2023

CITATION
Mari MA and Müller M-L (2023) Social cognition and Relevance: How stereotypes impact the processing of definite and indefinite descriptions. *Front. Commun.* 8:1088861. doi: 10.3389/fcomm.2023.1088861



# Social cognition and Relevance: How stereotypes impact the processing of definite and indefinite descriptions

Magali A. Mari* and Misha-Laura Müller

Institute of Communication and Cognitive Sciences, Cognitive Science Center, University of Neuchâtel, Neuchâtel, Switzerland

This paper focuses on the impact of social cognition on thes processing of linguistic information. More specifically, it brings some insights to Relevance theory's construal of Meaning$_{NN}$, which seeks to account for non-propositional meanings. It shows, through two experiments, how gender and nationality-related stereotypes guide the processing of definite and indefinite descriptions. Experiment 1 consists of a self-paced reading task (with 59 French native speakers), introducing information confirming vs. violating gender stereotypes within a nominal phrase (NP). The NP (e.g., "chirurgien/chirurgienne", "surgeon$_{male/female}$") was itself introduced either by a definite article (presupposition) or an indefinite article (assertion). Results showed that information violating gender stereotypes was costlier to process than stereotype-congruent information. Moreover, when information violated gender stereotypes, definite descriptions became significantly costlier than indefinite ones, because they required the identification of a salient referent which contradicted stereotypical expectations. Experiment 2 tested the effects of definite vs. indefinite NP on processing nationality-related stereotypes in a self-paced reading task (with 49 French native speakers). Participants read definite vs. indefinite NPs referring to representatives of a country. The NP was subsequently paired with information that confirmed vs. contradicted nationality stereotypes. Results showed that information contradicting nationality stereotypes were significantly costlier to process than information confirming stereotypes. Furthermore, when information contradicted nationality stereotypes, indefinite descriptions (which promote a single occurrence reading) failed to facilitate information processing compared to definite descriptions (which promote a generalized representation of the social category). Overall, the present findings are consistent with research on stereotypes, in that they show that stereotype-incongruent information affect sentence processing. Importantly, while Experiment 1 revealed that stereotypes affected the processing of linguistic markers, Experiment 2 suggested that linguistic markers could not modulate the processing of stereotypes.

KEYWORDS

**social cognition, gender stereotype, nationality stereotype, language processing, Relevance theory, massive modularity, definite description**

# 1. Introduction

This paper lies at the intersection of social cognition and pragmatics. Using tools from the study of stereotypes, it contributes to the theoretical framework of Relevance theory (Sperber and Wilson, 1986, 2015). Relevance theory argues that meaning derivation is guided by a comprehension heuristic. When exposed to an ostensive verbal stimulus, the listener seeks for optimal relevance, minimizing processing costs, to obtain most cognitive effects through the acquisition, reinforcement, or revision of a belief (cf. Sperber and Wilson, 2015, p. 135).

Initially, Relevance theory developed Grice's theory of implicatures (Grice, 1957, 1975), providing a cognitive explanation for pragmatic inferences responsible for explicit and implicit meanings. However, in recent years, it focused more on argumentation and literary studies (Sperber et al., 2010; Mercier and Sperber, 2011; Cave and Wilson, 2018; Wharton and Strey, 2019). In this context, Relevance theory presented a new research agenda oriented toward a broader approach to ostensive communication (Sperber and Wilson, 2015): it is emphasized that an adequate theory of meaning should include not only "determinate propositions" conveyed by linguistic stimuli, but also non-propositional meanings conveyed by verbal and non-verbal cues. Among the examples mentioned, they present the following exchange, for which the levels of analysis are broader than those initially proposed in classical approaches in pragmatics:

(1) Rob: Do you live in London?
Jen: I live in Chelsea

(Sperber and Wilson, 2015, p. 144)

In the above, Jen implicitly answers Rob's question in the affirmative, given that Chelsea is a neighborhood in London. However, the relevance of the utterance will not only depend on determinate contents (on the level of explicature or implicature), but also on less determinate ones, triggered by the tone of voice or the social status of the speakers (Sperber and Wilson, 2015, p. 144). Here, Sperber and Wilson point out that when social status is manifest, they will guide inferences in different ways. For example, depending on their respective social status, Jen's utterance may express closeness because she shares more specific information about where she lives, or it may express a sense of social superiority that can be paraphrased as "I don't live in just any part of London".

According to Relevance theory, social status corresponds to "encyclopedic information", such as gender or nationality stereotypes in this study. Encyclopedic information is used by addressees to construct the context which guides them in making interpretive inferences. In response to an ostensive stimulus, the recipient constructs contextual hypotheses on the basis of information that is more or less salient, and, respectively, more or less easy to process. The construction of the context will allow the addressee to infer the premises leading to the derivation of an intentional explicature or implicature intended by the speaker (Sperber and Wilson, 1986, p. 37).

Furthermore, it should be noted that the most recent lines of research in Relevance theory argue that the comprehension heuristic should be conceived as a broader process than initially defined, accounting for less determinate meanings and including non-verbal cues (Sperber and Wilson, 2015, p. 137). Following these new perspectives, Wilson (2016, p. 15) argues that linguistic markers may activate clusters of domain-specific modules of cognition, such as mindreading, emotion reading, or social cognition. The activation of these domain-specific modules is presumed to have an effect on the relevance-guided comprehension heuristic.

The present study aims to contribute to current discussions in Relevance theory by testing the impact of gender and nationality-related stereotypes on the processing of specific linguistic information, namely definite and indefinite descriptions.

## 1.1. The processing of stereotypes

While reading, one must not only visually process the written words but also understand their underlying meaning. To comprehend a text, readers draw on different sources of knowledge, namely linguistic, orthographic, and general world knowledge (Perfetti and Stafura, 2014; Kendeou et al., 2016). As pointed out by Relevance theory [cf. example (1)], making inferences lies at the core of comprehension. With respect to reading tasks: readers retrieve information from memory to construct a mental representation of a text (Graesser et al., 1994; Elbro and Buch-Iversen, 2013; Kendeou et al., 2016). The mental representation combines elements that are derived explicitly from the text, as well as elements that are implicit, coming from the readers' previously acquired knowledge (Gygax et al., 2021). As such, readers' world knowledge plays an essential role in reading comprehension.

While reading a word or a sentence, related concepts are automatically activated in semantic memory (Gerrig and McKoon, 1998; O'Brien et al., 1998; Rapp and van den Broek, 2005; Rubio-Fernández, 2013). For instance, upon reading sentence (2), concepts such as LAWYER*, LAW COURT*, or CRIMINAL* are likely to be activated and be more accessible in readers' memory:

(2) The judge sentenced a burglar to two years in prison.

Similarly, theoretical accounts of stereotyping propose that a given situation might increase the accessibility of stereotypic knowledge in memory (Gilbert and Hixon, 1991; Quadflieg and Macrae, 2011; Rees et al., 2020). For instance, upon reading sentence (2), stereotypical representations of "the judge" and "a burglar" will be activated in readers' memory, allowing readers to hold expectations about the likely traits, features, and behaviors of the two protagonists (Klein and Bernard, 2015; Beukeboom and Burgers, 2019). From this perspective, stereotypes function as heuristics as they guide expectations about members of a social category and are rapidly processed (Krieglmeyer and Sherman, 2012; Müller and Rothermund, 2014). When information violates a stereotype, more cognitive effort is required to access stereotype-incongruent information from associative memory, leading to increased processing difficulty (Banaji and Hardin, 1996; Kutas and Federmeie, 2000; Bartholow and Dickter, 2008; White et al., 2009). Importantly, previous research showed that regardless of personal opinions, people in the same context tend to be knowledgeable about the stereotypes in their culture (Devine, 1989; Lepore and Brown, 1997; Moskowitz et al., 1999; Quadflieg and Macrae, 2011;

Beukeboom and Burgers, 2019). As such, if a word or a sentence refers to a social category, readers within the same culture will spontaneously produce inferences about this social category and will most likely hold similar stereotypical expectations.

Overall, the effects of stereotype information on reading are well-documented. An important line of research assessed how gender stereotypes affect anaphora resolution of personal or reflexive pronouns (see for e.g., Carreiras et al., 1996; Kennison and Trofe, 2003; Duffy and Keir, 2004; Irmen, 2007; Esaulova et al., 2014; Reali et al., 2015). These studies showed that reading times of anaphoric pronouns were longer when stereotypical expectations about role nouns did not match the gender of the pronoun (e.g., "The *firefighter* burned *herself* while rescuing victims from the building", Duffy and Keir, 2004, p. 553). Another line of research tested whether readers make inferences about the gender of a person upon reading a role noun (and so, not only when required by the anaphora). For instance, Garnham et al. (2002) designed a study in which readers could make inferences about the gender of a character, without involving anaphora resolution (e.g., "The soldier drove to the playgroup after work, and picked up one of the children, who said 'Look what I did today daddy!'", Garnham et al., 2002, p. 442, see also Reynolds et al., 2006; Lassonde, 2015). Their findings suggest that readers automatically encode gender when they are exposed to role nouns, even though gender information is not crucial for comprehension (Gygax et al., 2021). Altogether, past studies revealed that, in various languages and cultures, reading is slowed down when *gender* is incongruent with stereotypical representations of role nouns (e.g., in English: Garnham et al., 2002; Reynolds et al., 2006; Lassonde, 2015; in Norwegian: Gabriel et al., 2017; in German: Irmen, 2007; Esaulova et al., 2014; in Italian: Cacciari et al., 1997; in Spanish: Carreiras et al., 1996). However, because gender is considered as a primary social category[1] (Brewer, 1988; Fiske, 1998), it is not clear whether information processing would be affected by stereotypical expectations about other social categories, that are less primary than gender, such as nationality stereotypes. For this reason, the present study compared, in a first experiment, the effects of well-studied primary stereotypes (i.e., professions associated with gender). In a second experiment, we assessed whether the observed effects also apply to less studied stereotypes, such as nationality stereotypes. Both experiments were designed to assess the extent to which stereotypes impact the processing of specific linguistic information (see next section below).

## 1.2. Research question and hypotheses

This study builds on Singh et al.'s (2016) experiment testing the impact of plausible vs. implausible contexts on the processing of definite and indefinite descriptions. While definite descriptions trigger a presupposition of a salient referent, indefinite descriptions merely introduce a new referent (Singh et al., 2016, p. 619, but see also Sperber and Wilson, 1986, p. 706).

According to Relevance theory, presuppositions and assertions can be distinguished in terms of foreground and background implications. While asserted contents contribute to relevance by providing additional cognitive effects, presuppositions contribute to relevance by saving efforts (Sperber and Wilson, 1986, p. 706). With respect to indefinite descriptions, they will be responsible for generating more effort because they present a noun as a new referent to the reader. This is not the case with definite descriptions, which present the noun phrase as "familiar" in context (cf. Heim, 1982; Roberts, 2003; Schwarz, 2009).

In Experiment 1, definite descriptions occur in a context that requires a bridging inference (Clark, 1975). That is to say that the referent is not explicitly mentioned in the preceding context, thereby requiring the construction of a link between the context (e.g., a hospital) and the noun (e.g., a/the surgeon). However, each context sentence was designed to have a strong semantic proximity with the target definite description, which facilitates processing (Haviland and Clark, 1974; Garrod and Sanford, 1977; Clifton, 2013; Schwarz, 2019). Bridging inferences are even easier in Experiment 2, as they involve a context introducing a superordinate concept (i.e., the name of a country), followed by a noun for a subordinate concept (i.e., the inhabitants of the country).

With regard to Singh et al.'s (2016) study, they hypothesized that implausible contexts, as in (3) below, would lead to an increased processing difficulty upon reading the following sentence. Moreover, within the implausible condition in (3), the definite description was expected to be significantly more difficult to process than the indefinite one, as it requires the identification of a salient referent in an incompatible context.

Singh et al. used two methods to test participants, namely a self-paced-reading task and a stop-making-sense task[2]. In both methods, participants read a plausible vs. implausible context sentence (3), followed by a definite or indefinite noun phrase (henceforth NP). Implausible contexts were expected to make the target NPs significantly costlier than plausible ones. Furthermore, as mentioned above, definite NPs like (3b) were expected to be costlier than indefinite ones like (3a) within implausible contexts:

(3) Mary went to the $\text{beach}_{\text{plausible}}$ / $\text{office}_{\text{implausible}}$ a few hours ago.
(3a) A lifeguard warned her there about the weather.
(3b) The lifeguard warned her there about the weather.

(Singh et al., 2016, p. 631)

Singh et al. observed an effect of context plausibility, where implausible contexts made the target NP (*A/The* lifeguard) significantly costlier to process than plausible ones. However, no significant difference was found when comparing definite NPs with indefinite ones. An effect was only found in the stop-making-sense task, when summing over all participants: In this

1 Gender is considered as a primary social category because attention to gender emerges early (see for e.g., Quinn et al., 2002) and because children of 3–4 years of age are already aware of conventional gender stereotypes (see for e.g., Weinraub et al., 1984; Leinbach et al., 1997; Shutts et al., 2009).

2 In the stop-making-sense task, participants were instructed to continue making words appear, segment by segment, as long as the sentences made sense. As soon as an incoming word or phrase did not make sense in the context of the preceding words/phrases, participants were asked to end the task (cf. Singh et al., 2016, p. 615).

case the proportion of dropouts was significantly higher in the presupposition condition (*The* lifeguard) than in the assertion condition (*A* lifeguard) (Singh et al., 2016, p. 617). Importantly, no effect was found between presupposition and assertion conditions in the self-paced reading task (Singh et al., 2016, p. 618). In a replication of Singh et al.s' study, using eye-tracking and self-paced-reading tasks (Müller and Mari, 2021), found significant results for plausibility effects, but no difference between definite and indefinite articles in the implausible condition, just like Singh et al.

The present study seeks to take these experiments further, using congruent vs. incongruent stereotypes instead of plausible vs. implausible contexts. The use of stereotypes, instead of context plausibility, is beneficial on two levels. First, it solves the problem of "context plausibility", which involves effects from various possible sources (e.g., surprise, comprehension problems, or also typicality effects). Importantly, the stimuli in this experiment used only plausible contexts, thus allowing the critical variable to be isolated, excluding surprise effects or problems attributable to the comprehension of the utterance. Second, as presented in the previous section, stereotypes are widely studied and well-understood in terms of reading tasks.

Experiment 1 consisted in a self-paced-reading task, assessing the impact of gender stereotypes (i.e., a primary social category) on the processing of asserted vs. presupposed contents. More specifically, Experiment 1 aimed to replicate previous findings on the effects of gender stereotypes on reading times cross-linguistically (with French speaking Swiss participants) and sought to identify the specific time course of processing gender stereotypic information. To this end, Experiment 1 tested the following hypothesis:

> Hypothesis 1: information violating gender stereotypes (4a) would be costlier to process than stereotype-congruent information (4b), within a compatible context (4).

(4) Lucienne est allée à l'hôpital le mois dernier. (Context sentence)
(4a) *La/Une chirurgienne* l'a opérée avec une grande précision. (Stereotype-incongruent)
(4b) *Le/Un chirurgien* l'a opérée avec une grande précision. (Stereotype-congruent)

> [Lucienne went to the hospital last month. (Context sentence)
> *The/A* $surgeon_{female}$ operated on her with great precision. (Stereotype-incongruent)
> *The/A* $surgeon_{male}$ operated on her with great precision. (Stereotype-congruent)]

Furthermore, and as in Singh et al. (2016) and Müller and Mari (2021), Experiment 1 tested whether definite NPs would lead to longer processing compared to indefinite NPs when the information contradicts a gender stereotype. In this case, the identification of a salient referent, required for definite NPs, is inconsistent with the encoding of stereotype-incongruent information.

> Hypothesis 2: stereotype-incongruent NPs would be costlier to process when presupposed through a definite description (e.g., "la chirurgienne"; "the $surgeon_{female}$") than when asserted through an indefinite description (e.g., "une chirurgienne"; "a $surgeon_{female}$").

Experiment 2 focused on the processing of nationality-related stereotypes, i.e., a secondary social category, and their interaction with definite and indefinite descriptions. To our knowledge, only two papers have studied the processing of secondary social categories. Dickinson (2011) focused on stereotypical inferences regarding heterosexuality during reading tasks, and Lassonde (2015) assessed stereotypical expectations regarding the behaviors of social groups[3]. Whereas, Lassonde (2015) found that reading times were longer for information that violated stereotypical expectations about social groups, Dickinson (2011) failed to reach conclusive results. Thus, given the limited information available on secondary social categories, it is worth providing new investigations.

In Experiment 2, participants first read a context sentence introducing the name of a country. Two countries were alternatively presented, for example Italy vs. Japan, as presented below (5). The second sentence introduced a redundant NP ("A/The Italian/s" vs. "A/The Japanese"), followed by an attribute ("great seducer/s") which was congruent (5a) or incongruent (5b) with a stereotype:

(5) Mathilde est allée en Italie/ au Japon le week-end dernier. (Context sentence)
(5a) Un/Les italien/s *a/ont joué au/x grand/s séducteur/s* durant tout le séjour. (Stereotype-congruent)
(5b) Un/Les japonais *a/ont joué au/x grand/s séducteur/s* durant tout le séjour. (Stereotype-incongruent)

> [Mathilde went to Italy/Japan last weekend. (Context sentence)
> An/The Italian/s *played the great seducer/s* during the whole stay. (Stereotype-congruent)
> A/The Japanese *played the great seducer/s* during the whole stay. (Stereotype-incongruent)]

Theoretical perspectives on stereotyping propose that any kind of stereotype-incongruent information should be difficult to process because it requires more cognitive effort to access this information from associative memory (see for e.g., Banaji and Hardin, 1996; Kutas and Federmeie, 2000; Bartholow and Dickter, 2008; White et al., 2009). Drawing from this perspective, the following hypothesis was tested:

3 In Lassonde's (2015) study, the stereotype-incongruent information was introduced by a whole sentence (e.g., "The nuns said there was not enough alcohol" vs. "The rockers said there was not enough alcohol", Lassonde, 2015, p. 161). In Dickinson's (2011) study, the stereotype-incongruent information was initiated by anaphora resolution (e.g., "Last night, in the packed movie theatre Hannah screamed loudly until her wife held her close", Dickinson, 2011, p. 457).

> Hypothesis 3: information violating nationality stereotypes should elicit longer reading times than stereotype-congruent information.

As illustrated above, the noun introducing the inhabitants of the country was preceded either by a plural definite or by an indefinite article. It should be noted that in French (in which language the study was conducted), plural definites invite a generic reading (Robinson, 2005, p. 18), thereby favoring a generalized and taxonomic representation of the social category described. However, in the present experimental setting, plural definites remain referential, thus fulfilling the condition of a presupposition (i.e., referring to a salient referent in the context)[4]. As for indefinite NPs, they favor a single occurrence reading, thus presenting information about the social category as singular in the provided context[5].

Following Sperber and Wilson (1986, p. 706), Experiment 2 tested whether readers would save processing efforts for presupposed contents, as opposed to asserted ones:

> Hypothesis 4: definite articles would be read more quickly than indefinite articles because they presuppose a referent which is highly salient (redundant in the context).

Finally, we conducted exploratory analyses to evaluate whether stereotype-incongruent information would be easier to process when introduced by an indefinite article (single occurrence reading) than by a definite one (generalized and definitional representation of the social category). These exploratory analyses aimed to evaluate whether stereotype-incongruent information was easier to process when it is under the scope of an indefinite description, as it promotes the reading of only one occurrence of an unexpected representation.

# 2. Experiment 1

Experiment 1 aimed to further assess the specific processing time course of gender stereotypes and to replicate previous findings (i.e., that gender stereotype-incongruent information is costly to process) cross-linguistically with French speaking Swiss participants (Hypothesis 1). Experiment 1 also investigated whether stereotype-incongruent information is costlier to process when it is presupposed through a definite description compared to when it is asserted though an indefinite description (Hypothesis 2).

## 2.1. Methods

### 2.1.1. Participants

For Experiment 1, 59 French speaking participants were recruited from a university in Switzerland. Only native French speakers were selected to participate in the experiment. The total sample size was set before data collection and based on the sample size estimation for "counterbalanced designs" developed by Westfall et al. (2014: 2026). The sample size estimation was conducted on Westfall and colleagues' website (https://jakewestfall.shinyapps.io/crossedpower/). We used the "standard case" values of variance components (VPCs; Westfall et al., 2014, p. 2025), with a power set at 0.85, a medium effect size of $d = 0.50$, and a number of 22 stimuli. The sample size estimation revealed that 58.8 participants were required. No additional participant was recruited once the pre-set sample size of 59 participants was reached. Following Singh et al. (2016) and Müller and Mari (2021), which employed the same experimental design as the current study, we excluded data from participants who had an accuracy rate for comprehension questions lower than 65%. This led to the exclusion of two participants. The final sample size resulted in 57 participants (31 women and 26 men; with an age mean of 23.87 years old, $SD = 4.29$).

### 2.1.2. Materials

The materials were constructed following a 2 × 2 design, manipulating (a) information about the social category, which either confirmed or violated stereotypical expectations, and (b) the NP introducing the social category, either with a definite article "le/la", "the" (presupposition condition), or with an indefinite article "un/une", "a/an" (assertion condition). The stimuli were created from the same model as those employed in Singh et al. (2016) and Müller and Mari (2021). Namely, the stimuli consisted in sets of two sentences written in French. The first sentence introduced a context, which was then followed by a target sentence matching or violating a gender stereotype. The target sentence introduced a specific agent marked grammatically by gender (e.g., chirurgien/chirurgienne, surgeon$_{male/female}$). The NP of the target sentence, i.e., the NP containing the social category concept, was introduced either with a definite article (working as a presupposition trigger) or an indefinite article (working as an assertion). In the end, each stimulus varied across four conditions which manipulated the effect of stereotypes and the article preceding the NP: (1) *stereotype-congruent and definite NP*, (2) *stereotype-congruent and indefinite NP*, (3) *stereotype-incongruent and definite NP* and (4) *stereotype-incongruent and indefinite NP* (see Table 1).

Gender stereotypes were based on a selection of role nouns tested in Misersky et al. (2014) as well as additional role nouns commonly found in French speaking Switzerland. A list of 50 role nouns were pre-tested on another sample of 36 subjects (50% self-identified as women) from the same population as the

4 Robinson (2005, p. 18) points out that generic readings in French can be encoded either by singular or plural definite descriptions. One test allowing to claim the presence of a generic reading is to see if the predicate cannot apply to an individual (*Paul est rare [*Paul is rare]). In the present experimental setting, predicates can apply to an individual (e.g., Paul a joué au grand séducteur [Paul *played the great seducer*.]). This speaks in favor of a non-generic reading of the stimuli.

5 Grice (1975, p. 56) provides examples with indefinite articles to illustrate the phenomenon of generalized conversational implicatures [e.g., "X is meeting a woman this evening."; "X went into a house yesterday and found a tortoise (…)"]. He explains that the use of the indefinite article promotes the inference that the item is unfamiliar. In the present experimental design, the use of the indefinite article includes the notion of unfamiliarity. However, it also promotes a single occurrence reading.

TABLE 1 Example of a stimulus of Experiment 1 in the four experimental conditions.

| Condition | Context sentence | Target sentence |
|---|---|---|
| Stereotype-congruent and definite NP | Lucienne \| est allée \| à l'hôpital \| le mois dernier. | Le chirurgien \| l'a opérée \| avec une grande précision. |
| | Lucienne \| went to \| the hospital \| last month. | The surgeon$_{male}$ \| operated on her \| with great precision. |
| Stereotype-congruent and indefinite NP | Lucienne \| est allée \| à l'hôpital \| le mois dernier. | Un chirurgien \| l'a opérée \| avec une grande précision. |
| | Lucienne \| went to \| the hospital \| last month. | A surgeon$_{male}$ \| operated on her \| with great precision. |
| Stereotype-incongruent and definite NP | Lucienne \| est allée \| à l'hôpital \| le mois dernier. | La chirurgienne \| l'a opérée \| avec une grande précision. |
| | Lucienne \| went to \| the hospital \| last month. | The surgeon$_{female}$ \| operated on her \| with great precision. |
| Stereotype-incongruent and indefinite NP | Lucienne \| est allée \| à l'hôpital \| le mois dernier. | Une chirurgienne \| l'a opérée \| avec une grande précision. |
| | Lucienne \| went to \| the hospital \| last month. | A surgeon$_{female}$ \| operated on her \| with great precision. |

Vertical bars (|) indicate the separation between each segment.

participants of Experiment 1. The pre-test was run on Qualtrics (Provo, UT) and followed a procedure similar to Misersky et al. (2014). Participants had to indicate on a 5-point Likert scale their opinion about the extent to which role nouns consisted of women or men[6]. Response options included "mostly women," "more women," "as much women as men," "more men," "mostly men" (coded as 1 for "mostly women" and 5 for "mostly men"). Role nouns that obtained the smallest scores ($M = 2.23$, $SD = 0.32$) were selected as female stereotypes and roles nouns that obtained highest scores ($M = 3.79$, $SD = 0.38$) were used as male stereotypes. In total, 22 stimuli were used, half related to female role nouns and half related to male role nouns. An additional set of 24 filler sentences was used to mask the purpose of the experiment. The complete list of stimuli and fillers is available at https://osf.io/b8h5q/.

Stimuli were also pre-tested in terms of plausibility. A total of 34 raters indicated, *via* Qualtrics (Provo, UT), the probability to encounter a specific social agent in a given situation (e.g., seeing surgeons in a hospital). The questions were asked in the following form: "*Si Marie va dans un hôpital, il est probable qu'elle rencontre … chirurgien.ne(s)*", "If Mary went to the hospital, it is likely that she encounters … surgeon(s)". Raters could choose between "zero," "only one," "one or more," "necessarily more than one" to replace the dots. For the selected stimuli, 76.4% of the raters chose "one or more"[7], assuring that the stimuli were considered as plausible.

### 2.1.3. Procedure

The experiment was created with and ran on E-Prime 2.0 software (Psychology Software Tools, Inc., 2012). We masked the purpose of the study from participants by informing them that they would participate in a study that investigated the links between causal information and its effects on the perception of narrativity in a reading task. Participants were instructed to read the sentences for comprehension. At the end of the study, the real purpose of the study was revealed.

Before running the experiment, participants were asked to indicate their age, gender, and mother tongue. The stimuli and fillers were then presented in sentence segments of 2–3 words (see Table 1), written in white 16-point Arial font on a black background. Each trial started with a white fixation cross on a black background, presented for 500 ms in the middle of the screen. The first segment then appeared on the screen. Participants would then press the spacebar to display the segments consecutively. This procedure prevented participants from displaying the whole sentence before reading it. Participants read only one condition of each stimulus, and as many stimuli from each of the four conditions, resulting in a within-subjects and within-stimuli design (Brauer and Curtin, 2018). Stimuli and fillers were presented randomly. Comprehension questions were used to assess whether participants remained attentive during the whole task. Comprehension questions were asked about the filler sentences only, and directly followed the corresponding filler. Participants answered yes or no by pressing on the "E" or "I" keys on the keyboard, according to the location of the yes/no answers on the screen. The experiment started with six practice trials, including one comprehension question, to familiarize participants with the task.

## 2.2. Results

### 2.2.1. Data analysis

The effects of stereotype-congruent vs. incongruent information and the article preceding the NP on information processing were measured by reading times, i.e., the time spent reading a sentence segment before clicking on the space bar to make a new segment appear. Three segments are considered for the analysis: (a) the critical segment consisting in the stereotype-congruent/incongruent information and the definite/indefinite NP, (b) the first spillover segment that follows the critical segment, and (c) the second spillover segment [see example (6); vertical bars separate the sentence segments]. The two segments following the critical segment are traditionally included in the analysis of self-paced reading measures. In this way, it is possible to assess

6 For example, *Veuillez indiquer si vous trouvez que plus de femmes ou d'hommes occupent la profession de chirurgiens/chirurgiennes,* [Please indicate whether you find that more women or men work as surgeons$_{male/female}$].

7 21.5% of the raters chose "necessarily more than one", 2.1% of the raters chose "only one", and none of the raters chose "zero".

potential processing difficulties that emerged or persisted after reading the critical segment (Liversedge et al., 1998).

(6) Lucienne | est allée | à l'hôpital | le mois dernier. | La chirurgienne $_{\text{critical segment}}$ | l'a opérée $_{\text{spillover 1}}$ | avec grande précision $_{\text{spillover 2}}$.
[Lucienne | went to | the hospital | last month. | The surgeon$_{\text{female critical segment}}$ | operated on her $_{\text{spillover 1}}$ | with great precision $_{\text{spillover 2}}$.]

Reading times below 100 ms and above 4,000 ms were excluded from the final dataset, leading to the suppression of 1.4% of data and a final dataset of 1,238 datapoints (the dataset is available at https://osf.io/b8h5q/). The data were logarithmically transformed to meet the assumptions of mixed effects model analyses (i.e., homoscedasticity, linearity, and normality). Data analysis was conducted on RStudio (R Core Team, 2019, version 3.6.0), using the lme4 package (Bates et al., 2015b).

### 2.2.2. Model selection

Model specification was driven by the experimental design, as recommended by experts in the field (Barr et al., 2013; Winter and Wieling, 2016; Brauer and Curtin, 2018). Fixed predictors are composed of the interaction between the stereotype condition (stereotype-congruent or incongruent information) and the NP condition (definite or indefinite article). Due to the repeated measures design, both subjects and stimuli created non-independence in the data and were thus included as by-subjects and by-stimuli random effects (Brauer and Curtin, 2018, p. 401). According to Barr et al. (2013), each fixed predictor that vary within-unit should include a random slope, as well as interactions when all factors vary within-units. In the present study, the stereotype condition and the NP condition varied both within-subjects and within-stimuli. Consequently, reading times were assessed with the following maximal mixed effect model: model <- lmer (log reading times ~ stereotype * NP + (stereotype + NP + stereotype*NP | subjects) + (stereotype + NP + stereotype*NP | stimuli).

The maximal mixed effect model for the three analyzed segments converged. For the first spillover segment convergence was reached by using the built-in optimization procedure "bobyqa" of the lme4 package (Bates et al., 2015b). This procedure has been acknowledged as one of the "remedies" that should be used to achieve convergence[8] (Brauer and Curtin, 2018, p. 404). The maximal mixed effect models for the three segments analyzed resulted however in a singular fit. Singular fits are indicators that the models are overparametrized and that they should be reduced to parsimonious models, balancing at the same time the Type I error rate and statistical power (Bates et al., 2015a,b; Matuschek et al., 2017). We thus conducted a random effect Principal Component Analysis, using the rePCA function of the lme4 package (Bates et al., 2015b). Goodness of fit was estimated with the likelihood ratio test (LRT) and AIC/BIC criteria (Bates et al., 2015a; Matuschek et al., 2017). The resulting models for reading times of the three segments are displayed in Table 2. The details of model selection and comparison are available at https://osf.io/b8h5q/. We also ran models including participants' gender to assess potential differences between self-identified male and female participants. For all three analyzed segments, we found no effect of gender. Gender was thus not included as a fixed predictor in the final models.

TABLE 2 Resulting parsimonious models for reading times on the three analyzed segments of Experiment 1.

| Segment analyzed | Final parsimonious model |
|---|---|
| Critical segment | lmer (log critical segment ~ stereotype * NP + (stereotype*NP \|\| subjects) + (stereotype \|\| stimuli)) |
| First spillover | lmer (log spillover1 ~ stereotype * NP + (stereotype + stereotype*NP \|\| subjects) + (stereotype \|\| stimuli)) |
| Second spillover | lmer (log spillover2 ~ stereotype * NP + (stereotype + NP \|\| subjects) + (stereotype \|\| stimuli)) |

Parsimonious models were selected after a random effect principal component analysis, estimation of goodness of fit with likelihood ratio test, AIC, and BIC criteria (Bates et al., 2015a,b; Matuschek et al., 2017). Details of model selection are available at https://osf.io/b8h5q/.

### 2.2.3. Reading times for the critical segment

The effect of stereotype-congruent and incongruent information on reading times was first assessed. The analysis revealed that there was no main effect of stereotype on reading times of the critical segment. Although reading times of stereotype-incongruent information ($M$ = 1,106.44 ms, $SD$ = 604.44) were longer than reading times of stereotype-congruent information ($M$ = 1,046.3 ms, $SD$ = 571.2), this difference was not significant, $t(115.5) = -1.64$, $p = 0.103$ (see Table 4).

When looking at the effect of definite and indefinite NPs only, we found no significant differences again between definite ($M$ = 1,067.1 ms, $SD$ = 589.7) and indefinite NPs ($M$ = 1,085.5 ms, $SD$ = 587.8), $t(1, 107) = 0.552$, $p = 0.581$ (see Table 4).

No interaction effect between stereotype information and the article preceding the NP were observed, $t(347.7) = 0.109$, $p = 0.913$ (see Table 4).

### 2.2.4. Reading times for the two spillover segments

The two segments following the critical segment were analyzed to assess whether a processing difficulty emerged after reading a particular segment (Liversedge et al., 1998).

The analysis revealed a main effect of stereotype information on reading times for the first spillover segment, $t(82.12) = -2.4$, $p = 0.019$. Reading times of stereotype-incongruent information ($M$ = 852.61 ms, $SD$ = 456.48) were significantly longer than reading times of stereotype-congruent information ($M$ = 812.06 ms, $SD$ = 420.55). These results support Hypothesis 1, namely that information violating gender stereotypes is costlier to process than stereotype-congruent information.

8 Failures of convergence are often due to the complexity of the random effect structure required by the experimental design. For the present study, the number of parameters estimates was 25, which might have been too high to reach a stable maximum likelihood estimation given the 1,238 datapoints (Barr et al., 2013; Bates et al., 2015a; Brauer and Curtin, 2018; Winter, 2019).

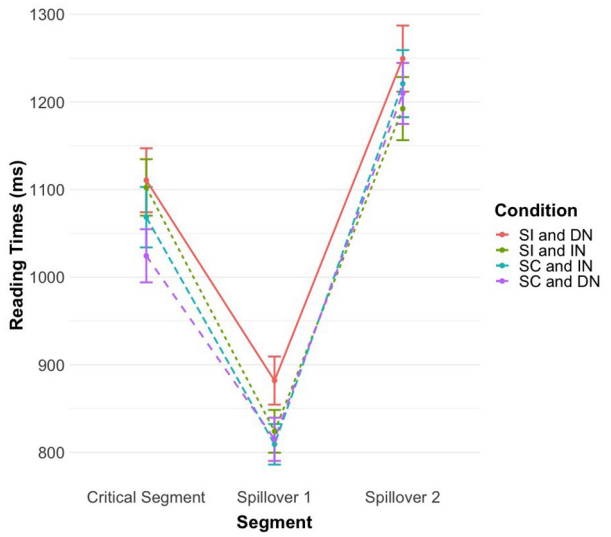



FIGURE 1
Mean reading times in milliseconds for Experiment 1. Mean and standard error reading times (raw data) for each segment in Experiment 1. SC, stereotype-congruent; SI, stereotype-incongruent; DN, definite NP; IN, indefinite NP.

A main effect of the article was also observed on the first spillover, with longer reading times after definite NPs ($M$ = 848.11 ms, $SD$ = 458.79) than after indefinite NPs ($M$ = 816.66 ms, $SD$ = 418.55), $t(1,076) = -2.18$, $p = 0.029$. No interaction effect between stereotype and NP was observed, $t(286.5) = 1.21$, $p = 0.229$ (see Table 4). Let us note that these results contradict the hypothesis of Relevance theory, namely that definite articles should be read more quickly than indefinite articles in plausible contexts. This issue is raised in the Section 2.3.

Contrast analyses were nonetheless conducted to assess whether within stereotype-incongruent conditions, longer processing times were observed with definite NPs as opposed to indefinite NPs (Hypothesis 2). These analyses revealed that reading times were significantly longer after reading stereotype-incongruent information introduced by a definite article ($M$ = 881.89 ms, $SD$ = 480.11) than when introduced by an indefinite article ($M$ = 824 ms, $SD$ = 430.98), $t(1,076) = 2.18$, $p = 0.029$ (see Figure 1). Moreover, reading an incongruent stereotype introduced by a definite article was significantly costlier than in any other condition (see Table 3).

The analysis of the second spillover segment revealed no effect of the stereotype information [$t(62.45) = -0.80$, $p = 0.429$], no effect of the NP [$t(178.65) = -1.33$, $p = 0.186$], and no interaction effect between stereotype and NP [$t(1,048) = 0.79$, $p = 0.428$] (see Table 4). These results suggest that the difficulty of processing emerged right after reading the critical segment and stopped immediately after the first spillover, namely, once the verb phrase was reached (see Figure 1).

## 2.3. Discussion

Experiment 1 replicated previous findings on the impact of gender stereotypes on processing times, using self-paced reading tasks. Unlike previous studies which analyzed reading times of complete sentences (e.g., Carreiras et al., 1996; Cacciari et al., 1997; Reynolds et al., 2006; Dickinson, 2011; Lassonde, 2015) or acceptability judgements (e.g., Garnham et al., 2002; Sato et al., 2013; Gabriel et al., 2017), this study presented sentence segments of 1–3 words, allowing a moment-by-moment analysis of processing difficulty. The analysis revealed that the processing difficulty of stereotype-incongruent role nouns was delayed to the first spillover segment. This is in line with eye-tracking studies showing that reading times are significantly slowed down upon and/or after reading a pronoun that led to a mismatch between a role noun and its anaphoric pronoun (e.g., reading "electrician" followed by "she," Reali et al., 2015, see also Kennison and Trofe, 2003; Duffy and Keir, 2004; Irmen, 2007; Esaulova et al., 2014). Experiment 1 thus replicates previous findings with French speaking Swiss participants: Information violating gender stereotypes is costlier to process than stereotype-congruent information.

Turning now to the effect of NPs, Experiment 1 revealed that definite NPs led to longer reading times than indefinite NPs on the spillover region. As noted above, these results contradict the assumption of Relevance theory, according to which definite articles should be less costly to process than indefinite articles within a plausible context. However, it should be noted that the observed longer processing time of definite articles was mainly driven by the processing of stereotype-incongruent information which generated a significant slowdown. Indeed, as revealed by contrast analyses, stereotype-incongruent information introduced by a definite NP (e.g., the surgeon$_{female}$) were significantly costlier than all other conditions. Within stereotype-congruent conditions (e.g., surgeon$_{male}$) definite articles were slightly (6 ms) costlier to process than indefinite articles. This is in line with Singh et al. (2016, p. 617), who also observed a slight slowdown with definite articles, as opposed to indefinite ones. It is likely that this experimental setup makes the processing of the definite articles costly, due to a difficulty to identify the referent in the previous context sentence. As we pointed out above (Section 1.2), the stimuli required a bridging inference, which is not necessary when the noun is preceded by an indefinite article, as it merely introduces a new referent.

Together these findings show that the processing of definite NPs, which requires the identification of a salient referent, is significantly affected by stereotypical representations. Normally, in plausible contexts, definite NPs should require little processing efforts (Sperber and Wilson, 1986, p. 706). However, the present experimental design suggests that the processing of definite descriptions interacts with social cognitive modules, generating a significant slowdown, despite a plausible context.

In sum, Hypothesis 1 was confirmed: Stereotypes are predictive of linguistic processing, where information incongruent with gender stereotypes is significantly costlier to process than stereotype-congruent information. Furthermore, Hypothesis 2 was also confirmed: Definite NPs were significantly costlier than indefinite NPs within the incongruent-stereotype condition. Regarding Hypothesis 2, it should be stressed that previous experiments on context plausibility (Singh et al., 2016; Müller and Mari, 2021) were not able to show a significant difference between definite and indefinite articles within implausible condition.

TABLE 3 Contrast analyses for the first spillover segment of Experiment 1.

| Conditions | M (*SD*) | SC and IN | SC and DN | SI and IN |
|---|---|---|---|---|
| | | *t*-test | | |
| SC and IN | 809 (406) | | | |
| SC and DN | 815 (435) | $t(85.9) = 0.43, p = 0.672$ | | |
| SI and IN | 824 (431) | $t(66.8) = 0.77, p = 0.447$ | $t(79.2) = 0.38, p = 0.703$ | |
| SI and DN | 882 (480) | $t(66.7) = 2.72, p = 0.008$ | $t(82.1) = 2.40, p = 0.019$ | $t(1,076) = 2.18, p = 0.029$ |

SC, stereotype-congruent; SI, stereotype-incongruent; DN, definite NP; IN, indefinite NP.

# 3. Experiment 2

Experiment 2 assessed (a) whether information violating expectations about secondary social categories (nationality stereotypes) are costly to process as is information violating gender stereotypes (Hypothesis 3), and (b) whether definite articles are more quickly read than indefinite articles in redundant contexts (Hypothesis 4). Exploratory analyses were conducted to evaluate whether stereotype-incongruent information is easier to process when introduced by an indefinite article (single occurrence reading) than by a definite one (generalized representation of the social category).

## 3.1. Methods

### 3.1.1. Participants

For Experiment 2, 49 French speaking participants were recruited from a university in Switzerland. As in Experiment 1, only native French speakers were selected to participate in the experiment. The total sample size was set before data collection and based on a sample size estimation as conducted for Experiment 1. Using the website of Westfall et al. (2014; https://jakewestfall.shinyapps.io/crossedpower/), we set the values for "counterbalanced designs" as in the two previous experiments, namely with the "standard case" values of VPCs, a power of 0.80, a medium effect size of $d = 0.50$ and a number of stimuli of 20. The sample size estimation revealed that 48.8 participants were required. No additional participant was recruited once the pre-set sample size of 49 participants was reached. Similar to Experiment 1, we controlled that participants provided a minimum of 65% accuracy rate for comprehension questions. All participants responded with more than 65% accuracy. The final sample size resulted in 49 participants (28 women and 21 men; with an age mean of 23.06 years old, $SD = 3.53$).

### 3.1.2. Materials

The stimuli were constructed in a similar way to Experiment 1. They consisted in two sentences written in French, with the first sentence introducing the context, and the following sentence matching or violating a nationality-related stereotypes. The target sentence introduced a social category, i.e., inhabitants of a country. The NP introducing the social category, was either a definite NP (working as a presupposition, favoring a generalized and taxonomic representation of the social category) or an indefinite NP (working as an assertion, favoring a single occurrence reading of the stereotype). As in Experiment 1, each stimulus varied across four conditions: (1) *stereotype-congruent with definite NP*, (2) *stereotype-congruent with indefinite NP*, (3) *stereotype-incongruent with definite NP*, and (4) *stereotype-incongruent with indefinite NP* (see Table 5).

Nationality-related stereotypes were based on folk stereotypes found in everyday speech (e.g., in movies, jokes, hearsay, comics, etc.) in the region of French speaking Switzerland. A list of 90 nationality stereotypes were pre-tested on another sample of 36 subjects (50% self-identified as women) from the same population as the final sample of Experiment 2. The pre-test was run on Qualtrics (Provo, UT) and asked participants to indicate on a 5-point Likert scale their opinion about diverse statements[9]. Response options included "agree," "somewhat agree," "neither agree nor disagree," "somewhat disagree," "disagree" (coded as 1 for "agree" and 5 for "disagree"). Statements that obtained the smallest scores ($M = 2.35$, $SD = 0.33$) were selected as nationality stereotypes and statements that obtained the highest scores ($M = 3.87$, $SD = 0.29$) were used as nationality counter-stereotype in the present study. In total, we used 20 stimuli, half matching nationality stereotype and half violating nationality stereotypes[10]. An additional set of 24 filler sentences was used to veil the purpose of the experiment. The complete list of stimuli and fillers is available at https://osf.io/b8h5q/.

### 3.1.3. Procedure

The procedure was the same as the one described in Experiment 1.

## 3.2. Results

### 3.2.1. Data analysis

As in Experiment 1, the effects of stereotype-congruent vs. incongruent information and the definite vs. indefinite article were measured by reading times. Four segments were considered

9 For example, *À quel point êtes-vous d'accord avec la proposition suivante: "les Japonais sont de grands séducteurs"* [To what extent do you agree with the following statement "the Japanese are great seducers"].

10 The plausibility was not pre-tested for Experiment 2, because the inhabitants introduced in the target sentence corresponded to those of the country presented in the context sentence (e.g., going to Japan and seeing Japanese is highly plausible).

| | Fixed effects | | | | | | Random effects | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | CI (95%) | *t*-value | DF | *p*-value | | Var. | SD |
| **Critical segment** | | | | | | | | | |
| (Intercept) | 6.874 | 0.049 | (6.78, 6.97) | 142.94 | 89.82 | <0.001 | Subjects intercept | 0.09 | 0.30 |
| Stereotype-congruent | −0.052 | 0.031 | (−0.11, 0.11) | −1.64 | 115.5 | 0.103 | Subjects slope (stereotype∗NP) | 0.01 | 0.09 |
| Indefinite NP | −0.017 | 0.030 | (−0.04, 0.08) | 0.55 | 1,107 | 0.581 | Stimuli intercept | 0.005 | 0.07 |
| Interaction | −0.005 | 0.045 | (−0.08, 0.09) | 0.11 | 347.7 | 0.913 | Stimuli slope (stereotype) | 0.001 | 0.04 |
| **First spillover** | | | | | | | | | |
| (Intercept) | 6.662 | 0.049 | (6.56, 6.76) | 137.03 | 76.99 | <0.001 | Subjects intercept | 0.11 | 0.33 |
| Stereotype-congruent | −0.065 | 0.027 | (−0.12, −0.01) | −2.40 | 82.12 | **0.019** | Subjects slope (stereotype) | 0.004 | 0.06 |
| Indefinite NP | −0.055 | 0.025 | (−0.10, −0.005) | −2.18 | 1,076 | **0.029** | Subjects slope (stereotype∗NP) | 0.003 | 0.05 |
| Interaction | 0.044 | 0.036 | (−0.03, 0.12) | 1.21 | 286.5 | 0.229 | Stimuli intercept | 0.003 | 0.06 |
| | | | | | | | Stimuli slope (stereotype) | 0.001 | 0.03 |
| **Second spillover** | | | | | | | | | |
| (Intercept) | 7.008 | 0.062 | (6.89, 7.13) | 113.52 | 62.41 | <0.001 | Subjects intercept | 0.09 | 0.31 |
| Stereotype-congruent | −0.026 | 0.033 | (−0.09, 0.04) | −0.80 | 62.45 | 0.429 | Subjects slope (stereotype) | 0.01 | 0.11 |
| Indefinite NP | −0.039 | 0.029 | (−0.09, 0.02) | −1.33 | 178.56 | 0.186 | Subjects slope (NP) | 0.007 | 0.09 |
| Interaction | 0.030 | 0.038 | (−0.04, 0.10) | 0.79 | 1,048 | 0.428 | Stimuli intercept | 0.04 | 0.20 |
| | | | | | | | Stimuli slope (stereotype) | 0.004 | 0.06 |

DF, degrees of freedom; SE, standard error; CI, confidence interval; Var., Variance; SD, standard deviation. Values in bold are significant at $p < 0.05$ (calculated using Satterthwaites approximations). The selected mixed effects models are presented in Table 2.

TABLE 5 Example of a stimulus of Experiment 2 in the four experimental conditions.

| Condition | Context sentence | Target sentence |
|---|---|---|
| Stereotype-congruent and definite NP | Mathilde \| est allée \| en Italie \| le week-end dernier. | Les Italiens \| ont joué \| aux grands séducteurs \| durant tout le séjour. |
| | Mathilde \| went to \| Italy \| last weekend. | The Italians \| played \| the great seducers \| during the whole stay. |
| Stereotype-congruent and indefinite NP | Mathilde \| est allée \| en Italie \| le week-end dernier. | Un Italien \| a joué \| au grand séducteur \| durant tout le séjour. |
| | Mathilde \| went to \| Italy \| last weekend. | An Italian \| played \| the great seducer \| during the whole stay. |
| Stereotype-incongruent and definite NP | Mathilde \| est allée \| au Japon \| le week-end dernier. | Les Japonais \| ont joué \| aux grands séducteurs \| durant tout le séjour. |
| | Mathilde \| went to \| Japan \| last weekend. | The Japanese \| played \| the great seducers \| during the whole stay. |
| Stereotype-incongruent and indefinite NP | Mathilde \| est allée \| au Japon \| le week-end dernier. | Un Japonais \| a joué \| au grand séducteur \| durant tout le séjour. |
| | Mathilde \| went to \| Japan \| last weekend. | A Japanese \| played \| the great seducer \| during the whole stay. |

Vertical bars (|) mark presentation boundaries (i.e., sentences segments).

for the analysis: (a) the one containing the definite/indefinite NP that introduced the social category (i.e., inhabitants of a country), (b) the spillover segment to assess potential persistence of processing difficulty, (c) the segment presenting stereotype-congruent/incongruent information, and (d) its spillover segment [see example (7); vertical bars separate the sentence segments]:

(7) Mathilde | est allée | au Japon | le week-end dernier. | Les Japonais$_{\text{critical segment 1}}$ | ont joué$_{\text{spillover 1}}$ | aux grands séducteurs$_{\text{critical segment 2}}$ | durant tout le séjour $_{\text{spillover 2}}$.
[Mathilde | went to | Japan | last weekend. | The Japanse$_{\text{critical segment 1}}$ | played$_{\text{spillover 1}}$ | the great seducers$_{\text{critical segment 2}}$ | during the whole stay$_{\text{spillover 2}}$.]

Similar to Experiment 1, reading times below 100 ms and above 4,000 ms have been excluded from the final dataset. This data exclusion resulted in the suppression of 1.1% of data and a final dataset of 969 datapoints (the dataset is available at https://osf.io/b8h5q/). The data were logarithmically transformed to meet the assumptions of mixed effects model analyses and data analysis was conducted on Rstudio (R Core Team, 2019, version 3.6.0), using the lme4 package (Bates et al., 2015b).

### 3.2.2. Model selection

We followed the same procedure as in Experiment 1 to specify the model (i.e., model selection based on the experimental design). The first segment under investigation in the present experiment did not mix the types of NPs and stereotype information. As illustrated in example (7), the first segment varies only in terms of the article used, namely definite or indefinite. The information violating/confirming stereotypes is only introduced in the seventh segment (critical segment 2). As a consequence, reading times on the first critical segment and the first spillover were assessed with the following maximal mixed effect model: model1 <- lmer (log reading times ~ NP + (NP | subjects) + (NP | stimuli). On the other hand, reading times of the second critical segment and its spillover could be affected by both the type of NP and stereotype information. Therefore, reading times of those remaining segments were analyzed with the following maximal mixed effect model: model2 <- lmer (log reading times ~ stereotype * NP + (stereotype*NP | subjects) + (stereotype*NP | stimuli).

TABLE 6 Resulting parsimonious models for reading times on the three analyzed segments of Experiment 2.

| Segment analyzed | Final parsimonious model |
|---|---|
| Critical segment 1 | lmer (log critical segment 1 ~ NP + (NP \|\| subjects) + (0 + NP \|\| stimuli)) |
| Spillover 1 | lmer (log spillover1 ~ NP + (NP \|\| subjects) + (1 \| stimuli)) |
| Critical segment 2 | lmer (log critical segment 2 ~ stereotype * NP + (1 \| subjects) + (stereotype + NP \|\| stimuli)) |
| Spillover 2 | lmer (log spillover 2 ~ stereotype * NP + (NP \|\| subjects) + (NP \|\| stimuli)) |

Parsimonious models were selected after a random effect Principal Component Analysis, estimation of goodness of fit with likelihood ratio test, AIC, and BIC criteria (Bates et al., 2015a,b; Matuschek et al., 2017). Details of model selection are available at https://osf.io/b8h5q/.

The maximal mixed effect model for the four analyzed segments reached convergence. For the two critical segments and the two spillover segments, the built-in optimization procedures "nlminbwrap" and "bobyqa" of the lme4 package (Bates et al., 2015b) were used, respectively. The maximal mixed effect models for the four segments resulted however in a singular fit, indicating that the models were overparametrized. We thus conducted a random effect Principal Component Analysis, using the rePCA function of the lme4 package (Bates et al., 2015b). Goodness of fit was estimated with the likelihood ratio test (LRT) and AIC/BIC criteria (Bates et al., 2015a; Matuschek et al., 2017). The resulting models for reading times of the four segments are displayed in Table 6. The details of model selection and comparison are available at https://osf.io/b8h5q/.

### 3.2.3. Reading times for the first critical and spillover segments

The analysis revealed that there was no main effect of definite or indefinite NPs on reading times of the critical segment, $t(44.23) = 1.59$, $p = 0.118$. The next segment was also analyzed to assess if a processing difficulty emerged after reading the definite/indefinite NPs. The analyses revealed a significant difference between reading

times, $t(51.24) = 2.49$, $p = 0.016$, with longer reading times following an indefinite NP ($M = 884.69$ ms, $SD = 493.92$) compared to a definite NP ($M = 825.08$ ms, $SD = 439.37$). This result confirms Hypothesis 4, where definite NPs were expected to be read more quickly than indefinite ones (see Table 7). These results are compatible with Relevance theory, which argues that definite descriptions allow to spare cognitive efforts (compared with indefinite descriptions).

### 3.2.4. Reading times for the second critical and spillover segments

The analysis revealed that there was a main effect of stereotype information on reading times of the second critical segment. Reading times of stereotype-incongruent information ($M = 1{,}309.26$ ms, $SD = 676.07$) were significantly longer than reading times of stereotype-congruent information ($M = 1{,}199.43$ ms, $SD = 610.02$), $t(51.6) = -3.05$, $p = 0.004$. A main effect of stereotype was also observed for the spillover segment, with longer reading times for the stereotype-incongruent condition ($M = 1{,}066.05$ ms, $SD = 538.86$) compared to the stereotype-congruent condition ($M = 993.24$ ms, $SD = 523.52$), $t(831.7) = -3.04$, $p = 0.002$ (see Figure 2). These results support Hypothesis 3, according to which stereotype-incongruent information elicit longer reading times than stereotype-congruent information. Importantly, these results provide evidence for the persistence of stereotype effects with secondary social categories.

No effect of definite/indefinite NPs and no interaction effect were observed for both the second critical segment and its spillover (see Table 7).

Contrast analyses were nonetheless conducted on both segments to explore the possibility that stereotype-incongruent information might be easier to process when introduced by an indefinite article as opposed to a definite article. These analyses revealed that reading times in the *stereotype-incongruent with indefinite NP* condition were not significantly faster than in the *stereotype-incongruent with definite NP* condition, $t(64.7) = 0.27$, $p = 0.786$ (critical segment 2) and $t(49.7) = 0.43$, $p = 0.666$ (second spillover). In other words, the processing of stereotype-incongruent information does not appear to be affected by linguistic markers of definiteness (see Figure 2).

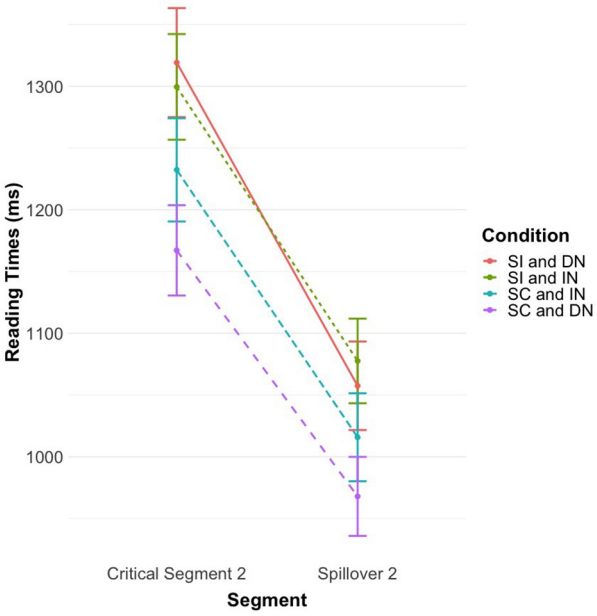



FIGURE 2
Mean reading times in milliseconds for Experiment 2 (second critical segment and spillover). Mean and standard error reading times (raw data) in Experiment 2 for the second critical segment and its spillover. SC, stereotype-congruent; SI, stereotype-incongruent; DN, definite NP; IN, indefinite NP.

## 3.3. Discussion

Experiment 2 revealed that stereotype-incongruent information about nationalities is longer to process than stereotype-congruent information. This finding provides further evidence that information confirming stereotypical expectations is easily processed, whereas information violating stereotypical expectations about secondary social categories is difficult to process (Hypothesis 3). Interestingly, the effects of information confirming/violating nationality stereotype already appeared on the critical segment and persisted in the spillover segment.

Regarding the effects of definite/indefinite NPs, Experiment 2 tested whether indefinite descriptions were costlier to process than definite descriptions, as proposed by Relevance theory. The present study confirmed Hypothesis 4, showing that definite NPs led to faster reading times than indefinite ones. This finding is all the more interesting in light of Experiment 1, where the fast reading of definite descriptions was disrupted because of information violating gender stereotypes. Finally, we explored whether indefinite articles (i.e., the representation of a single occurrence within a kind) could facilitate the processing of stereotype-incongruent information. These analyses revealed that indefinite articles could not make stereotype-incongruent information easier to process than when subjected to a generalization (i.e., a plural definite). Thus, in this experimental setup, the processing effects of stereotypes appear to be stronger than linguistic markers.

# 4. General discussion

The present study investigated the effects of social modules of cognition on the relevance-guided comprehension heuristic across two experiments, in order to shed light on the relevance comprehension heuristic. Both experiments assessed the extent to which stereotypes impact the processing of specific linguistic information. Experiment 1 aimed to replicate previous findings on the effects of gender stereotypes on reading cross-linguistically. Experiment 2 sought to investigate the effects of secondary social categories, i.e., nationalities.

The results of Experiment 1 showed that information violating gender stereotypes is longer to process than stereotype-congruent information (Hypothesis 1). This finding goes in line with previous studies that investigated, in different languages, the effect of gender

TABLE 7 Statistical results of the selected parsimonious models for Experiment 2.

| | Fixed effects | | | | | | Random effects | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | CI (95%) | *t*-value | DF | *p*-value | | Var. | SD |
| **Critical segment 1** | | | | | | | | | |
| (Intercept) | 6.783 | 0.044 | (6.69, 6.87) | 152.78 | 52.98 | <0.001 | Subjects intercept | 0.08 | 0.29 |
| Indefinite NP | 0.050 | 0.031 | (−0.01, 0.11) | 1.59 | 44.23 | 0.118 | Subjects slope (NP) | 0.02 | 0.13 |
| | | | | | | | Stimuli (NP) | 0.002 | 0.05 |
| **Spillover 1** | | | | | | | | | |
| (Intercept) | 6.600 | 0.051 | (6.50, 6.70) | 129.10 | 55.97 | <0.001 | Subjects intercept | 0.11 | 0.33 |
| Indefinite NP | 0.060 | 0.024 | (0.01, 0.11) | 2.49 | 51.24 | **0.016** | Subjects slope (NP) | 0.006 | 0.07 |
| | | | | | | | Stimuli intercept | 0.002 | 0.05 |
| **Critical segment 2** | | | | | | | | | |
| (Intercept) | 7.063 | 0.059 | (6.95, 7.18) | 119.07 | 71.81 | <0.001 | Subjects intercept | 0.10 | 0.32 |
| Stereotype-congruent | −0.103 | 0.034 | (−0.17, −0.03) | −3.05 | 51.61 | **0.004** | Stimuli intercept | 0.02 | 0.14 |
| Indefinite NP | −0.008 | 0.031 | (−0.07, 0.05) | −0.27 | 69.63 | 0.787 | Stimuli slope (stereotype) | 0.005 | 0.07 |
| Interaction | 0.053 | 0.042 | (−0.03, 0.14) | 1.26 | 864.35 | 0.207 | Stimuli slope (NP) | 0.002 | 0.04 |
| **Spillover 2** | | | | | | | | | |
| (Intercept) | 6.864 | 0.051 | (6.76, 6.97) | 133.79 | 70.98 | <0.001 | Subjects intercept | 0.09 | 0.31 |
| Stereotype-congruent | −0.085 | 0.028 | (−0.14, −0.03) | −3.04 | 831.65 | **0.002** | Subjects slope (NP) | 0.008 | 0.09 |
| Indefinite NP | 0.015 | 0.033 | (−0.05, 0.08) | 0.45 | 67.86 | 0.654 | Stimuli intercept | 0.006 | 0.08 |
| Interaction | 0.026 | 0.039 | (−0.05, 0.10) | 0.648 | 831.09 | 0.517 | Stimuli slope (NP) | 0.09 | 0.31 |

DF, degrees of freedom; SE, standard error; CI, confidence interval; Var., Variance; SD, standard deviation. Values in bold are significant at $p < 0.05$ (calculated using Satterthwaites approximations). The selected mixed effects models are presented in Table 6.





10.3389/fcomm.2023.1088861

stereotypes on reading and anaphora resolution (cf. Section 1.1). Our findings with French speaking Swiss participants support the cross-linguistic evidence that gender is rapidly encoded during reading (Garnham et al., 2002; Gygax et al., 2021) and affects processing depending on whether the information communicated matches one's stereotypical expectations.

Furthermore, Experiment 1 revealed that stereotype-incongruent information makes the processing of presuppositional contents (definite articles) significantly costlier than assertions (indefinite descriptions) (Hypothesis 2). This is because the identification of a salient referent (required for definite NPs) is inconsistent with the encoding of stereotype-incongruent information. Importantly, these findings offer promising opportunities for the study of the relevance comprehension heuristic. While previous studies failed to reach conclusive results when using a broad category of plausibility (Singh et al., 2016; Müller and Mari, 2021), the present study revealed that the processing of definite descriptions is affected by stereotype information, generating a significant slowdown in narrowly defined plausible contexts. We suggest that the "plausibility of contexts", used in these previous studies, conflated different variables, such as surprise effects, comprehension problems as well as typicality effects.

In Experiment 2, we further assessed the effect of stereotypes about secondary social categories on processing, and revealed that information violating nationality stereotypes was costly to process (Hypothesis 3). This is consistent with Lassonde's (2015) study, showing that sentences containing stereotype-incongruent information about diverse social categories (e.g., nuns, rockstars) are costly to process. Together, these findings have some interesting implications, suggesting that any kind of stereotype-incongruent information would be difficult to process because more cognitive effort is required to access the information from associative memory. Unfortunately, this possibility has rarely been addressed, as most studies to date have focused only on gender stereotypes. In an endeavor to determine whether the effects of stereotypes on processing are consistent, future studies should investigate, with varying methodologies (e.g., response times, self-paced reading, eye-tracking, or event-related brain potentials) and across cultures, how and whether information processing is similarly affected by stereotype about various social categories.

Experiment 2 also revealed that definite descriptions are less costly than indefinite descriptions when the context is redundant. These results align with Relevance theory, which argues that definite descriptions allow to spare cognitive efforts (compared with indefinite descriptions). Moreover, our exploratory analysis showed that when a single occurrence was encoded linguistically (e.g., "A Japanese played the great seducer", as opposed to "The Japanese played the great seducers"), it did not facilitate the processing of incongruent-stereotype information. This suggests that the processing of incongruent-stereotype information cannot be modulated by linguistic markers[11].

Before concluding, let us note that the present study's findings bear some important considerations. Given the current context and issues, it seems particularly important to study how stereotypical information is processed. For instance, although much effort and attention has been paid to gender equality in the 21st century (e.g., the increasing use of inclusive language, the promotion of STEM professions among girls and women, or strikes for women's right), we still observed that some conceptions of gender roles remain unchanged. Moreover, current crises (namely the COVID-19 pandemic, the war in Ukraine, or the climate crisis) led people to rapidly form negative stereotypes about inhabitants of certain countries. By documenting how people process stereotypes during reading, the present study showed that information confirming a stereotype is easily processed and thus, might not be questioned or noticed, while still being significant in the relevance comprehension heuristic. Furthermore, this could play a role in the maintenance of stereotypical expectations and the emergence of prejudices.

We also stress two important limitations to the current study. First, it does not allow to make direct comparisons between the two tested stereotypes. Indeed, gender stereotype in/congruence occurred within the NP (e.g., "une chirurgienne"; "a surgeon$_{female}$") whereas nationality stereotype in/congruence stood in the relation between the NP and the predicate (e.g., "A/The Japanese *played the great seducer/s*"). In a future study, it would be worth testing these two stereotypes, and others, in a comparable way. Furthermore, our study focused only on narrow linguistic phenomena (definite vs. indefinite descriptions). While this may be an asset experimentally (limiting other variables weighing in the processing speed), further studies are needed to see if stereotypes also constrain the processing of other linguistic markers, such as other presupposition triggers.

Overall, the present study's findings suggest that stereotypes bring significant constraints on the processing of linguistic information. These elements are of interest for Relevance theory, insofar as they confirm that the comprehension heuristic is constrained by information which goes beyond propositional cues, such as the listener's knowledge about social categories. Across both experiments, stereotype-incongruent information was less salient than stereotype-congruent ones, making it cognitively more costly to process. Moreover, these findings suggest a possible hierarchy between social and linguistic information in the derivation of meaning: Indeed, while stereotype-incongruent information slows down the processing of definite descriptions (which are normally processed quickly), we did not observe that the processing of stereotype-incongruent information was facilitated when preceded by an indefinite article (single occurrence reading).

## Data availability statement

The original contributions presented in the study are available at https://osf.io/b8h5q/, further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by Commission d'Éthique de la Recherche (CER),

11 Let us note that an anonymous reviewer drew our attention to the potential problems of confounding variables in the experimental setup. For this reason, further experiments with a setup that better isolates the variables should be conducted to test this hypothesis.

Université de Neuchâtel. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

Data analysis, reporting, and data interpretation was conducted by MAM. All authors equally developed the study concept and contributed to the study design. Data collection was performed by all authors. The first draft of the manuscript was written by all authors, and all authors reviewed and revised it together. All authors approved the final version of the manuscript for submission.

## Funding

This research was funded by the University of Neuchâtel.

## 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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References


Banaji, M. R., and Hardin, C. D. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x

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

Bartholow, B. D., and Dickter, C. L. (2008). A response conflict account of the effects of stereotypes on racial categorization. *Soc. Cogn.* 26, 314–332. doi: 10.1521/soco.2008.26.3.314

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

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using Lme4. *J. Statist. Softw.* 67, 1–51. doi: 10.18637/jss.v067.i01

Beukeboom, C. J., and Burgers, C. (2019). How stereotypes are shared through language: a review and introduction of the social categories and stereotypes communication (SCSC) framework. *Rev. Commun. Res.* 7, 1–37. doi: 10.12840/issn.2255-4165.017

Brauer, M., and Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: a unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychol. Methods* 23, 389–411. doi: 10.1037/met0000159

Brewer, M. B. (1988). "A dual process model of impression formation," in *Advances in Social Cognition*, eds R. S. Wyer, and T. K. Srull (Hillsdale: Lawrence Erlbaum Associates), 1–36.

Cacciari, C., Carreiras, M., and Cionini, C. B. (1997). When words have two genders: anaphor resolution for italian functionally ambiguous words. *J. Mem. Lang.* 37, 517–532. doi: 10.1006/jmla.1997.2528

Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol. Hum. Exp. Psychol.* 49, 639–663. doi: 10.1080/713755647

Cave, T., and Wilson, D. (2018). *Reading Beyond the Code: Literature and Relevance Theory*. Oxford: Oxford University Press.

Clark, H. H. (1975). "Bridging," in *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing – TINLAP '75* (Morristown, NJ: Association for Computational Linguistics), 169–174. doi: 10.3115/980190.980237

Clifton, C. (2013). Situational context affects definiteness preferences: accommodation of presuppositions. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 487–501. doi: 10.1037/a0028975

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18. doi: 10.1037/0022-3514.56.1.5

Dickinson, J. (2011). The impact of 'violating the heterosexual norm' on reading speed and accuracy. *Psychology* 2, 456–459. doi: 10.4236/psych.2011.25071

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Mem. Cognit.* 32, 551–559. doi: 10.3758/BF03195846

Elbro, C., and Buch-Iversen, I. (2013). Activation of background knowledge for inference making: effects on reading comprehension. *Sci. Stud. Read.* 17, 435–452. doi: 10.1080/10888438.2013.774005

Esaulova, Y., Reali, C., and Stockhausen, L., v. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Lang. Cogn. Neurosci.* 29, 781–803. doi: 10.1080/01690965.2013.794295

Fiske, S. T. (1998). "Stereotyping, prejudice, and discrimination," in *The Handbook of Social Psychology*, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (Boston, MA: McGraw-Hill), 357–411.

Gabriel, U., Behne, D. M., and Gygax, P. M. (2017). Speech vs. reading comprehension: an explorative study of gender representations in Norwegian. *J. Cogn. Psychol.* 29, 795–808. doi: 10.1080/20445911.2017.1326923

Garnham, A., Oakhill, J., and Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Mem. Cogn.* 30, 439–446. doi: 10.3758/BF03194944

Garrod, S., and Sanford, A. (1977). Interpreting anaphoric relations: the integration of semantic information while reading. *J. Verbal Learning Verbal Behav.* 16, 77–90. doi: 10.1016/S0022-5371(77)80009-1

Gerrig, R. J., and McKoon, G. (1998). The readiness is all: the functionality of memory-based text processing. *Discour. Process.* 26, 67–86. doi: 10.1080/01638539809545039

Gilbert, D. T., and Hixon, J. G. (1991). The trouble of thinking: activation and application of stereotypic beliefs. *J. Pers. Soc. Psychol.* 60, 509–517. doi: 10.1037/0022-3514.60.4.509

Graesser, A. C., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychol. Rev.* 101, 371–395. doi: 10.1037/0033-295X.101.3.371

Grice, H. P. (1957). Meaning. *Philos. Rev.* 377–388. doi: 10.2307/2182440

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

Gygax, P., Sato, S., Öttl, A., and Gabriel, U. (2021). The masculine form in grammatically gendered languages and its multiple interpretations: a challenge for our cognitive system. *Lang. Sci.* 83, 101328. doi: 10.1016/j.langsci.2020.101328

Haviland, S. E., and Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *J. Verb. Learn. Verb. Behav.* 13, 512–521. doi: 10.1016/S0022-5371(74)80003-4

Heim, I. (1982). *The semantics of definite and indefinite noun phrases* (Ph. D. Dissertation). University of Massachusetts, Amherst, MA, United States.

Irmen, L. (2007). What's in a (Role) name? Formal and conceptual aspects of comprehending personal nouns. *J. Psycholinguist. Res.* 36, 431–456. doi: 10.1007/s10936-007-9053-z

Kendeou, P., McMaster, K. L., and Christ, T. J. (2016). Reading comprehension: core components and processes. *Policy Insights Behav. Brain Sci.* 3, 62–69. doi: 10.1177/2372732215624707

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

Klein, O., and Bernard, P. (2015). "Stereotypes in social psychology," in *International Encyclopedia of the Social & Behavioral Sciences*, ed J. D. Wright (Amsterdam: Elsevier), 446–452. doi: 10.1016/B978-0-08-097086-8.24097-X

Krieglmeyer, R., and Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *J. Pers. Soc. Psychol.* 103, 205–224. doi: 10.1037/a0028764

Kutas, M., and Federmeie, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trend. Cogn. Sci.* 4, 463–470. doi: 10.1016/S1364-6613(00)01560-6

Lassonde, K. A. (2015). Reducing the impact of stereotypical knowledge during reading. *Discourse Process.* 52, 149–171. doi: 10.1080/0163853X.2014.917221

Leinbach, M. D., Hort, B. E., and Fagot, B. I. (1997). Bears are for boys: metaphorical associations in young children's gender stereotypes. *Cogn. Dev.* 12, 107–130. doi: 10.1016/S0885-2014(97)90032-0

Lepore, L., and Brown, R. (1997). Category and stereotype activation: is prejudice inevitable? *J. Person. Soc. Psychol.* 72, 275–287. doi: 10.1037/0022-3514.72.2.275

Liversedge, S. P., Paterson, K. B., and Pickering, M. J. (1998). *Eye Movements and Measures of Reading Time. Eye Guidance in Reading and Scene Perception*. Amsterdam: Elsevier.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001

Mercier, H., and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behav. Brain Sci.* 34, 57–74. doi: 10.1017/S0140525X10000968

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., et al. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behav. Res. Methods* 46, 841–871. doi: 10.3758/s13428-013-0409-z

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., and Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *J. Pers. Soc. Psychol.* 77, 167–184. doi: 10.1037/0022-3514.77.1.167

Müller, F., and Rothermund, K. (2014). What does it take to activate stereotypes? Simple primes don't seem enough. *Soc. Psychol.* 45, 187–193. doi: 10.1027/1864-9335/a000183

Müller, M.-L., and Mari, M. A. (2021). Definite descriptions in the light of the comprehension vs. acceptance distinction: comparing self-paced reading with eye-tracking measures. *Front. Commun.* 6, 1–17. doi: 10.3389/fcomm.2021.634362

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., and Halleran, J. G. (1998). Updating a situation model: a memory-based text processing view. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1200–1210. doi: 10.1037/0278-7393.24.5.1200

Perfetti, C., and Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Sci. Stud. Read.* 18, 22–37. doi: 10.1080/10888438.2013.827687

Quadflieg, S., and Macrae, C. N. (2011). Stereotypes and stereotyping: what's the brain got to do with It? *Eur. Rev. Soc. Psychol.* 22, 215–273. doi: 10.1080/10463283.2011.627998

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., and Pascalis, O. (2002). representation of the gender of human faces by infants: a preference for female. *Perception* 31, 1109–1121. doi: 10.1068/p3331

Rapp, D. N., and van den Broek, P (2005). dynamic text comprehension: an integrative view of eeading. *Curr. Dir. Psychol. Sci.* 14, 276–279. doi: 10.1111/j.0963-7214.2005.00380.x

Reali, C., Esaulova, Y., Öttl, A., and Stockhausen, L. V. (2015). Role descriptions induce gender mismatch effects in eye movements during reading. *Front. Psychol.* 6, 1–13. doi: 10.3389/fpsyg.2015.01607

Rees, H. R., Ma, D. S., and Sherman, J. W. (2020). Examining the relationships among categorization, stereotype activation, and stereotype application. *Pers. Soc. Psychol. Bull.* 46, 499–513. doi: 10.1177/0146167219861431

Reynolds, D., Garnham, A., and Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *Q. J. Exp. Psychol.* 59, 886–903. doi: 10.1080/02724980543000088

Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguist. Philos.* 26, 287–350. doi: 10.1023/A:1024157132393

Robinson, H. M. (2005). *Unexpected (in) Definiteness: Plural Generic Expressions in Romance.* New Brunswick, NJ: The State University of New Jersey.

Rubio-Fernández, P. (2013). Associative and inferential processes in pragmatic enrichment: the case of emergent properties. *Lang. Cogn. Process.* 28, 723–745. doi: 10.1080/01690965.2012.659264

Sato, S., Gygax, P. M., and Gabriel, U. (2013). Gender inferences: grammatical features and their impact on the representation of gender in bilinguals. *Biling.: Lang. Cogn.* 16, 792–807. doi: 10.1017/S1366728912000739

Schwarz, F. (2009). *Two Types of Definites in Natural Language* (Ph. D. Dissertion). University of Massachusetts Amherst, Massachusetts, United States. Available online at: https://scholarworks.umass.edu/open_access_dissertations/122 (accessed February 20, 2023).

Schwarz, F. (2019). "Definites, domain restriction, and discourse structure in online processing," in *Grammatical Approaches to Language Processing*, eds C. J. Clifton, K. Carlson, and J. D. Fodor (Berlin: Springer), 187–208. doi: 10.1007/978-3-030-01563-3_10

Shutts, K., Banaji, M. R., and Spelke, E. S. (2009). Social categories guide young children's preferences for novel objects. *Dev. Sci.* 13, 599–610. doi: 10.1111/j.1467-7687.2009.00913.x

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

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., et al. (2010). Epistemic vigilance. *Mind Lang.* 25, 359–393. doi: 10.1111/j.1468-0017.2010.01394.x

Sperber, D., and Wilson, D. (1986). *Relevance: Communication and Cognition*. Vol. 142. Cambridge, MA: Harvard University Press.

Sperber, D., and Wilson, D. (2015). Beyond speaker's meaning. *Croat. J. Philos.* 15, 117–149.

Weinraub, M., Clemens, L. P., Sockloff, A., Ethridge, T., Gracely, E., and Myers, B. (1984). The development of sex role stereotypes in the third year: relationships to gender labeling, gender identity, sex-types toy preference, and family characteristics. *Child Dev.* 55, 1493–1503. doi: 10.2307/1130019

Westfall, J., Kenny, D. A., and Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen.* 143, 2020–2045. doi: 10.1037/xge0000014

Wharton, T., and Strey, C. (2019). "Slave of the passions: making emotions relevant," in *Relevance, Pragmatics and Interpretation*, *1st Edn*, eds K. Scott, B. Clark, and R. Carston (Cambridge University Press), 253–266. doi: 10.1017/9781108290593.022

White, K. R., Crites, S. L., Taylor, J. H., and Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Soc. Cogn. Affect. Neurosci.* 4, 191–98. doi: 10.1093/scan/nsp004

Wilson, D. (2016). Reassessing the conceptual–procedural distinction. *Lingua* 175, 5–19. doi: 10.1016/j.lingua.2015.12.005

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York, NY: Routledge.

Winter, B., and Wieling, M. (2016). How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *J. Lang. Evol.* 1, 7–18. doi: 10.1093/jole/lzv003

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 15 March 2023
DOI 10.3389/fpsyg.2023.1135129



OPEN ACCESS

EDITED BY
Caroline Jagoe,
Trinity College Dublin,
Ireland

REVIEWED BY
Kirsi Neuvonen,
University of Helsinki,
Finland
Maor Zeev-Wolf,
Ben-Gurion University of the Negev,
Israel

*CORRESPONDENCE
Camilo R. Ronderos
✉ camilorr@uio.no

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

RECEIVED 31 December 2022
ACCEPTED 21 February 2023
PUBLISHED 15 March 2023

CITATION
Ronderos CR and Falkum IL (2023) Suppression of literal meaning in single and extended metaphors.
*Front. Psychol.* 14:1135129.
doi: 10.3389/fpsyg.2023.1135129



# Suppression of literal meaning in single and extended metaphors

Camilo R. Ronderos* and Ingrid Lossius Falkum

Department of Philosophy, Classics, History of Art and Ideas, University of Oslo, Oslo, Norway

Within Relevance Theory, it has been suggested that extended metaphors might be processed differently relative to single metaphoric uses. While single metaphors are hypothesized to be understood *via* the creation of an *ad hoc* concept, extended metaphors have been claimed to require a switch to a secondary processing mode, which gives greater prominence to the literal meaning. Initial experimental evidence has supported a distinction by showing differences in reading times between single and extended metaphors. However, beyond potential differences in comprehension speed, Robyn Carston's 'lingering of the literal' account seems to predict qualitative differences in the interpretative mechanisms involved. In the present work, we test the hypothesis that during processing of extended metaphors, the mechanisms of enhancement and suppression of activation levels of literal-related features operate differently relative to single metaphors. We base our work on a study by Paula Rubio-Fernández, which showed that processing single metaphors involves suppressing features related exclusively to the literal meaning of the metaphoric vehicle after 1000 milliseconds of encountering the metaphor. Our goal was to investigate whether suppression is also involved in the comprehension of extended metaphors, or whether the 'lingering of the literal' leads to continued activation of literal-related features, as we take Carston's account to predict. We replicate existing results, in as much as we find that activation levels of literal-related features are reduced after 1000 milliseconds. Critically, we also show that the pattern of suppression does not hold for extended metaphors, for which literal-related features remain activated after 1000 milliseconds. We see our results as providing support for Carston's view that extended metaphor processing involves a prominent role of literal meaning, contributing towards explicating the links between theoretical predictions within Relevance Theory and online sentence processing.

KEYWORDS

**relevance theory, extended metaphors, metaphor comprehension, experimental pragmatics, figurative language**

## Introduction

The cognitive mechanisms responsible for metaphor comprehension have been the focus of much research throughout the last several decades (for reviews see Holyoak and Stamenković, 2018; Pouscoulous and Dulcinati, 2019). One reason for the sustained interest in this line of work is the apparent change in meaning that words undergo when used metaphorically. Take example (1):

> *1. John doesn't like physical contact, and even his girlfriend finds it difficult to come close to him. She feels rejected by his distant attitude every time he sees her. John is a cactus.*

It is clear that the word *cactus* is not used to refer to a type of plant, but to John's distant demeanor. But how does a comprehender go from understanding *cactus* as a plant to understanding it as a personality trait? Two related theories provide an answer this question by viewing metaphor comprehension as a form of category extension: Glucksberg's Dual Reference Account (Glucksberg, 2001, 2008), and Sperber and Wilson's Deflationary Account (Sperber and Wilson, 2008), which is embedded in the larger framework of Relevance Theory (Wilson and Sperber, 2012). Broadly speaking, both accounts state that comprehenders understand nominal metaphors as (1) by inferentially adjusting the meaning of the metaphoric vehicle (*cactus*) on the basis of the salient interpretative dimensions provided by the metaphoric topic (*John*), and crucially, encyclopedic information associated with the metaphor vehicle (*cactus*) together with the relevant context. Once this occurs, the topic is understood as being a member of an occasion-specific (*ad hoc*) category represented by the vehicle (McGlone and Manfredi, 2001; Glucksberg, 2003; Rubio Fernández, 2007; Sperber and Wilson, 2008). This idea borrows from previous work on the comprehension of *ad hoc* categories, which suggests that people are in general quite good at picking out the potential members of a newly created group (e.g., *things to bring to a picnic*, Barsalou, 1983). In this way, understanding so-called nominal metaphors ['A is a B' constructions such as (1)] is similar to understanding category inclusion statements such as *Papaya is a fruit*.

The relevance-theoretic view extends beyond the analysis of simple nominal metaphors. The view is that metaphors in general are understood on the basis of the same interpretative mechanism as other forms of lexical interpretation (hence the *deflationary* character of the account). According to the theory, lexical interpretation typically involves modulation of encoded word meanings, where *ad hoc* concepts are constructed in accordance with the hearer's occasion-specific expectations of relevance, based on the encoded concepts, a set of associated encyclopedic assumptions, and information derived from the utterance context (Wilson and Carston, 2007). *Ad hoc* concepts can either be more specific ('narrower') or more general ('broader') than the word's encoded meaning (as it is assumed to be stored in the mental lexicon). Critically, metaphors are said to result in both narrowing and broadening of the encoded meaning (Wilson and Carston, 2006; Carston, 2010). For example, in (1), *cactus* is broader than the encoded meaning because it includes a type of 'prickly' people, which the encoded concept excludes. It is also narrower than the encoded meaning because it excludes cacti without spikes (e.g., the spike-less peyote plant).

Broadening and narrowing can be thought of in terms of property promotion and demotion (Carston, 2002). According to Carston (2002), mentally stored concepts provide a memory link to three types of information: logical content, encyclopedic content, and lexical properties. Logical content is meaning-constituent (e.g., Cactus is a kind of plant), whereas encyclopedic content represents general world knowledge we associate with a specific concept (e.g., Cacti typically have spikes, they grow in the desert, etc.). During lexical modulation, some properties associated with the encoded concept are promoted whereas others are demoted. Property promotion and demotion can be conceptualized in psychological terms as the degree of activation of a particular property: A promoted property is highly activated, whereas a demoted property is not (e.g., Rubio Fernández, 2007). This would mean that, when constructing the (metaphorical) *ad hoc* concept CACTUS*,[1] certain encyclopedic features that are associated with the encoded meaning of *cactus* become highly activated (e.g., the fact that cacti have spikes), whereas those that are not relevant for the construction of the *ad hoc* concept have a substantially lower degree of activation (e.g., that cacti are a kind of plant).

Narrowing and broadening in the form of construction of *ad hoc* concepts are the outcomes of the interpretative process (Wilson and Carston, 2006; Carston, 2010). However, thinking of them as the degree of activation of encyclopedic features provides a link to the cognitive mechanisms potentially involved in metaphor comprehension, such as the mechanisms responsible for the suppression and enhancement of activation levels. Gernsbacher and Faust (1991) state that language comprehension in general is enabled by the enhancement and suppression of the activation levels of memory nodes. In this view, enhancement regulates the increase of activation of relevant information and suppression regulates the reduction of activation of irrelevant information. This led Gernsbacher et al. (2001) and, subsequently, Rubio Fernández (2007), to derive explicit hypotheses for category extension theories in terms of suppression and enhancement of associated features during metaphor comprehension. When processing a metaphor such as (1)–once the interpretative dimensions are made salient by the context and the metaphoric topic (that John is human and that his personality is being discussed)–comprehenders adjust the lexically encoded meaning of *cactus*. They do this by suppressing the activation of features that mismatch these dimensions (and are thus irrelevant for the unfolding interpretation, e.g., that cacti are plants), and by enhancing the activation of those that match (and are thus relevant for the interpretation, e.g., that cacti are prickly).

To test these claims, Gernsbacher et al. (2001) showed participants prime sentences that were either literal or metaphoric [*That large hammerhead is a shark* (literal), *that defense lawyer is a shark* (metaphoric)]. Participants then read and verified sentences that included words representing properties that were relevant or irrelevant for the *ad hoc* category [*sharks are tenacious* (relevant), *sharks are good swimmers* (irrelevant)]. The results showed that participants were faster to verify sentences about a metaphor-relevant feature following the metaphoric prime compared to when the sentence followed a literal prime. Conversely, verifying sentences describing a metaphor-irrelevant property was less costly when these followed a literal prime than when they followed a metaphoric prime.

Rubio Fernández (2007) provided further evidence using a cross-modal priming paradigm. In it, participants first heard a novel metaphor (*John is a cactus*) and, immediately after hearing the vehicle, read a target word and performed a lexical decision task. Critically, target words were shown at three possible intervals (0, 400 and 1,000 milliseconds after the end of the metaphor). Target words were either 'literal' superordinates of the metaphoric vehicle (and therefore irrelevant for the construction of the *ad hoc* category, e.g., *plant*) or distinctive properties at the core of the metaphoric meaning (relevant for constructing the *ad hoc* category, e.g., *spike*). The results showed that in the two earliest intervals (0 ms and 400 milliseconds), superordinates and distinctive properties were similarly activated.

1 We follow the convention within Relevance Theory to refer to *ad hoc* concepts *via* capital letters and an asterisk.

However, in the last interval (1,000 milliseconds), only distinctive properties remained active, while superordinates appeared to be suppressed. The results of both of these studies suggest that, during metaphor comprehension, metaphor-related features (i.e., the distinctive features) are suppressed while literal-related ones (i.e., the superordinates) remain active [for at least 1,000 milliseconds after the metaphor has been understood, according to Rubio Fernández (2007)]. This supports the *ad hoc* concept account by showing how comprehending a metaphor brings about the modulation of the encoded meaning of the metaphoric vehicle.

Despite the above-mentioned evidence, the issues of the mechanisms involved during comprehension and the validity of the *ad hoc* concept account are far from settled. This is in part due to the existence of rivaling theories that also have some empirical support (e.g., the structure mapping view of Gentner and Bowdle, 2008). It is also due to existing experimental evidence being rather limited in scope. The experiments discussed thus far focused on so-called single nominal metaphors (of the 'X is a Y' type). However, metaphors can come in all shapes and sizes. They can be expressed through verbs (*The sunflower danced in the sun*) or adjectives (*Miguel has a colorful personality*), for example. Examining a wide variety of cases, as argued by Holyoak and Stamenković (2018), is essential for assessing the robustness of a theory whose goal should be to account for the mechanisms behind metaphor processing independently of the morphosyntactic properties of the metaphoric vehicle. Steps have been taken in this direction, with various studies investigating the processing of non-nominal metaphors in recent years (e.g., Cardillo et al., 2012; Ronderos et al., 2021, 2022).

Besides the focus on nominal metaphors, an overwhelming majority of studies have looked at 'single' metaphors only, i.e., metaphors with a unique metaphoric vehicle. As a contrast to this, consider the case in which example (1) is slightly modified into (2):

> *2. John doesn't like physical contact, and even his girlfriend finds it difficult to come close to him. She feels pricked by his thorny attitude every time he sees her. John is a cactus*

In (2), the words *thorny* and *pricked* denote properties that are associated with the concept encoded by the word *cactus*, and thus their literal meanings are semantically related. Importantly, these words are used metaphorically in a way consistent with the metaphor in the sentence *John is a cactus*. As a whole, this passage constitutes what is known as an extended metaphor (Carston, 2010; Rubio-Fernández et al., 2016).

Carston (2010) suggests that extended metaphors might pose a problem for the relevance-theoretic analysis developed examining single metaphors only. This is because the mechanism proposed by Sperber and Wilson's deflationary account of metaphor in terms of *ad hoc* concept construction is a form of local meaning adjustment: each time a metaphoric vehicle is encountered (e.g., *cactus*), an *ad hoc* concept is created that differs from the encoded meaning in that it has been broadened (and typically also narrowed) (e.g., creating the *ad hoc* category CACTUS*). For an extended metaphor such as (2), this means that, upon encountering the words *thorny*, *pricked* and *cactus*, comprehenders would have to suppress literal features irrelevant to the metaphoric meaning each time. This would occur despite the fact that these three words are clearly related to each other and their literal meaning is likely to be highly activated given backwards and forward priming. The local lexical adjustment mechanism would result, according to Carston (2010), in a demanding and effortful process (but see Wilson, 2018, for a different perspective on this issue). Instead, Carston (2010) suggests, comprehenders might begin to maintain–through metarepresentation–the literal meaning of extended metaphors as a whole (because of how the literal meaning of the different vehicles 'lingers' and is therefore highly activated) and subject this literal interpretation to slower, broader inferences after the entire expression has been understood. In terms of online metaphor processing, the account proposed by Carston (2010) could be said to make one general prediction: metaphoric vehicles comprehended as part of extended metaphoric passages should be processed differently than the same vehicles encountered as stand-alone metaphors. There is some evidence to this effect that pre-dates Carston's 'lingering of the literal' account. Keysar et al. (2000) had participants read vignettes that included multiple metaphoric vehicles stemming from the same conceptual domain. They found that when the metaphors were novel (Experiment 2), target metaphoric sentences were read faster when preceded by related metaphoric vehicles relative to when the previous sentences contained no metaphors whatsoever. A similar result using the same paradigm was also reported by Thibodeau and Durgin (2008).

In a more explicit test of the 'lingering' account, Rubio-Fernández et al. (2016) used self-paced reading (Experiment 1), eye-tracking during reading (Experiment 2) and cued recall (Experiment 3) to examine processing differences between single and extended metaphors as well as literal controls. They found that participants took longer to read single metaphors relative to extended metaphors and literal controls (Experiment 1). In Experiment 2, they found that extended metaphors and literal controls were read similarly fast in an early reading measure (i.e., first-pass reading), whereas single metaphors were found to take longer to read. However, the difference between single and extended metaphors seemed to dissipate in later reading measures (i.e., total reading times). The authors interpret this as supporting Carston's view of a distinction between the two types of metaphor: Extended metaphors are first read as fast as a literal utterance (thus suggesting an early advantage in comprehension time for extended relative to single metaphors), and the late derivation of inferences in the extended metaphors case results in more effortful processing of extended vs. single metaphors in the latest moments of processing.

Despite the fact that this pattern of findings suggests a difference between both types of metaphors, it is unclear whether this difference is a *qualitative* or a *quantitative* one. In other words, it could be that extended metaphors are subjected to the same mechanisms as single metaphors but simply undergo the process of lexical modulation faster because of low-level priming brought on by the previously understood metaphors. This is akin to the view put forth by Wilson (2018), p. 195, who claims that differences between single and extended metaphors has more to do with a 'lingering of linguistic form' than a 'lingering of literal meaning', where the accumulation of metaphorical vehicles with related encoded meanings "will encourage some [hearers] to pay more attention to the exact wording of the [utterance] and search for further implications activated by the encoded meaning." Though Wilson (2018) does not explicate her view of the 'lingering of the linguistic form' in processing terms, one could explain the faster processing observed for extended compared to single metaphors as resulting not from a qualitative difference in processing between the two types of metaphor but from low-level semantic priming stemming from

processing various related metaphoric vehicles in a row [e.g., *pricked* and *thorny* in (1)]. This priming facilitates access to the entry in the mental lexicon of the subsequent related vehicle (*cactus*). Once the lexical entry has been accessed, processing continues normally, with comprehenders creating a new *ad hoc* category (CACTUS*). This would amount to a difference in *degree* of activation relative to single metaphors, and not a difference in kind. Extended metaphors would be comprehended faster than single metaphors but making use of the same mechanisms.

Alternatively, it could be that processing differences between extended and single metaphors are truly due to qualitative differences in the underlying mechanisms, as suggested by Carston (2010). In Carston's view, the persistent high activation of the closely related literal meanings of the metaphoric vehicles makes the creation of *ad hoc* categories for every single one of them too effortful. Instead, comprehenders metarepresent the literal meaning of the expressions throughout the processing of the extended metaphor. In processing terms, this would not only lead to differences in comprehension speed, but should also result in the involvement of different comprehension mechanisms: Single metaphors are processed *via* the construction of *ad hoc* categories (following the standard relevance-theoretic account), while extended metaphors are processed literally, with a literal representation of the entire passage being maintained and metarepresented even as metaphoric inferences are drawn. However, it is not entirely clear whether Carston's account would actually predict faster processing of extended metaphors, as suggested by Rubio-Fernández et al. (2016), or in *slower* processing due to metarepresentation of the literal meaning and derivation of a range of weak implicatures.

The goal of the current work is to test these two alternatives by examining the role of enhancement and suppression of the activation levels of literal features during processing of single and extended metaphors. As previously mentioned, others have suggested that when understanding a metaphor, features related exclusively to the literal meaning of a vehicle (what we refer to as 'literal-related features') are suppressed, while those related to the resulting metaphoric meaning (which we dub 'metaphor-related features') enhanced (Gernsbacher et al., 2001; Rubio Fernández, 2007). How should enhancement and suppression play out during processing of extended metaphors? One possibility is that the differences in comprehension effort for single relative to extended metaphors reported by Keysar et al. (2000), Thibodeau and Durgin (2008), and Rubio-Fernández et al. (2016) result in baseline differences in activation levels for both literal-related and metaphor-related features: being exposed to related metaphors facilitates lexical access to the subsequent related metaphoric vehicle, and therefore the recognition of both types of features is made easier at all time intervals. This would suggest that suppression and enhancement operate in basically the same way for extended metaphors as they do for single metaphors. They simply operate faster, in line with the view that one difference between the two types of metaphors is that extended metaphors, but not single ones, involve low-level priming of linguistic form. Another possibility would be that that there are differences in how suppression and enhancement unfold over time: While literal-related features are suppressed after around 1,000 milliseconds and metaphor-related features remain active in the case of single metaphors (Rubio Fernández, 2007), it could be that the mechanism of suppression is suspended when processing extended metaphors. This would result in sustained activation for literal-related features at different time intervals after processing the metaphor. This process would be in line with the view that literal meaning is metarepresented during the comprehension of extended metaphors (Carston, 2010).

To be clear, both alternatives are in principle compatible with a processing speed advantage for extended relative to single metaphors: The 'lingering of linguistic form' can be interpreted as a low-level priming effect that facilitates the retrieval of subsequent related metaphoric vehicles, whereas the 'lingering of the literal' leads the expression as a whole to be initially processed literally, without engaging in the construction of *ad hoc* concepts. However, it seems that only Carston's view would predict qualitative differences between single and extended metaphors in how suppression and enhancement of activation levels of literal features unfold over time. If the literal meanings of the metaphorical vehicles are metarepresented throughout the processing of the extended metaphor, it is likely that also features related to these literal meanings (and which are irrelevant to the metaphorical meanings) retain a high activation level, or at least are not suppressed to the same extent as if *ad hoc* concepts were created for each of the metaphorical vehicles. To test this key difference, we adapted Rubio Fernández (2007) seminal paradigm to a web-based experiment, and present the results of our study in the following section.

# Method

## Participants

We recruited a total of 460 participants *via* the online recruitment platform Prolific. Participants were all monolingual native speakers of American English between the ages of 18 and 35. They all had an IP-address from the United States during time of testing and reported being right-handed. Of these, 3 were excluded because of a technical problem. Of the remaining 457, 47 were excluded for not meeting the minimum accuracy requirement (i.e., achieving at least 70% accuracy in the lexical decision task across critical and filler trials). This left the total number of participants at 410.

## Materials and design

The starting point of our investigation was the experiment conducted by Rubio Fernández (2007). Since we intended to adapt the original experiment to a web-based task, we made three main adjustments. First, instead of using a cross-modal paradigm (where the prime is heard by participants and the target sentence read on the screen), we presented both primes and targets visually. This was done given the reduced amount of experimental control in a web-based experiment. For example, it was not possible for us to know if participants would use headphones or speakers or if they would be listening to music while completing the task. Therefore, presenting both prime and target in the written form seemed like an appropriate way to reduce noise and make sure that they were both understood. Second, we chose to use only a subset of the items used by Rubio Fernández (2007) to keep the experiment as short as possible and thus maximize the quality of the data collected from the online participants, following recommendations by Futrell (2012). We used 8 of the critical items from Rubio Fernández (2007) as our target items, and another

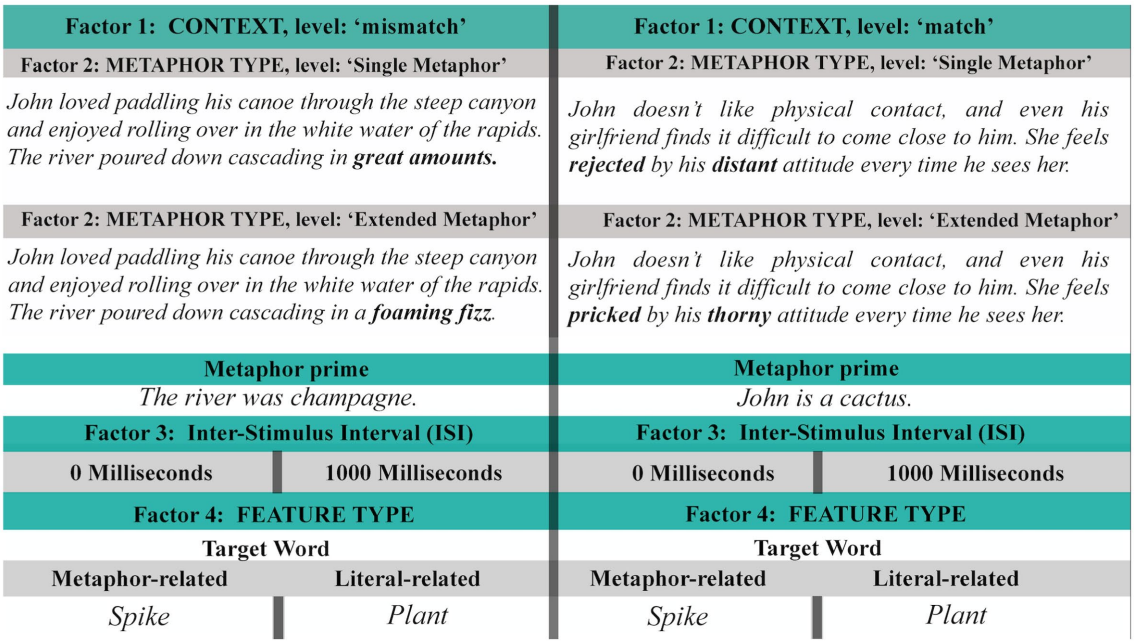



FIGURE 1
Example of a critical trial in all conditions.

group of 8 as fillers. Third, we reduced the number of Inter-Stimulus Intervals (ISI) tested relative to the original experiment from three (0, 400 and 1,000 milliseconds) to two (0 and 1,000 milliseconds). This was done to keep the number of experimental conditions to a minimum.

After making these adjustments we adapted the materials in order to create an extended-metaphoric version of each item. To do this, we added an additional context sentence prior to the nominal metaphor. For extended metaphors, this context sentence included additional metaphoric vehicles that drew from the same conceptual domain as the nominal metaphor. For single metaphors, the additional context sentence was a literal equivalent. This is exemplified in sentences (1) and (2), with all conditions reproduced in Figure 1 below.

The experiment was programmed using the PCIbex experimental software (Zehr and Schwarz, 2018) and had a 2X2X2X2 design (as seen in Figure 1): Participants first read a context (in a single or extended metaphor set up) and a metaphor prime that was either related or unrelated to the target word they would see afterwards (factor 1: CONTEXT, levels: 'match' vs. 'mismatch'). This factor ensures that we have a baseline measure of the lexical decision time for each word in the 'match' conditions: Every target word appeared following a single and extended matching or mismatching metaphor. In other words, responses to target words appeared in the absence of any potential semantic priming (i.e., when the metaphor prime is completely unrelated to the target, in the 'mismatch' level), as well as following a corresponding related single or extended metaphor (i.e., when the metaphor prime is critically related to the target, in the 'match' level). Each critical item was paired with another one to create the 'mismatch' conditions, so that every target word and every metaphor appeared equally in 'match' and 'mismatch' conditions across lists. The metaphor primes read were either instances of single or extended metaphors, depending on critical words being altered accordingly in the context (factor 2: METAPHOR TYPE, levels: 'single metaphor' vs. 'extended metaphor', see words in bold in Figure 1). After reading the metaphors, participants were forced to wait either 1,000 ms. Or to directly continue to the lexical decision task (factor 3: ISI, levels: 0 and 1,000 milliseconds). Finally, participants read the target words and performed a lexical decision task. Target words were either related to the irrelevant encoded literal meaning of a metaphoric vehicle only, or to the vehicle's relevant metaphoric meaning (factor 4: FEATURE TYPE, levels: 'literal-related vs. 'metaphor-related, or they were irrelevant to both, as in the 'mismatch' condition). As mentioned, we used an additional 8 of the original critical items of Rubio Fernández (2007) as fillers. These consisted of metaphoric primes and plausible English pseudo-words as targets. The pseudo-words were created using the online pseudo-word generator Wuggy[2], designed for use in psycholinguistic experiments. Both filler and critical trials in the experiment were metaphors, in line with the set-up used by Rubio Fernández (2007). Half of the filler items were extended metaphors and half were single metaphors. Half of the target pseudo-words were presented with an ISI of 1,000 milliseconds, with the other half having an ISI of 0 milliseconds. We created 16 experimental lists using a latin-squared design, distributing conditions in a balanced way across lists. However, since we only had 8 critical items, it was not possible for one participant to see all combinations of conditions in a single experimental list, given that this would have required at least 16 items. Instead, each participant saw 8 critical items in 8 different conditions, ensuring that they saw each level of each factor at least 4 times (across items), with all conditions evenly distributed through the 16 lists. This made our design a between-subjects one regarding the four-way

2 http://crr.ugent.be/programs-data/wuggy

interaction only, and a within-subjects design regarding all other comparisons. The master list including all 16 list combinations, together with all critical and filler items as well as the data and analysis script can be found on the project's OSF page: https://osf.io/eayj7/.

## Procedure

Participants first read the instructions of the experiment. They were then asked to make sure they were in a quiet environment, and were instructed to position their index fingers above the F and J key, and leave their thumbs over the SPACEBAR. In each trial, participants first read the context sentences. They had to press the SPACEBAR to continue once they finished reading what was presented to them. At this point, they were presented with the metaphor prime, which was always a nominal metaphor of the form 'X is a Y'. The metaphor prime was presented on screen until participants pressed the SPACEBAR, which they were told to press as soon as they had finished reading. Participants had to wait at least 1,500 milliseconds before being allowed to continue, and were required to press the SPACEBAR before 4,000 milliseconds after the prime metaphor was presented. If it took them more than 4,000 milliseconds to read the metaphor prime, the trial automatically exited and a text appeared on screen prompting them to be faster, without any data for this particular trial being recorded. This was done to discourage participants from reading the metaphors and waiting too long before moving onto the lexical decision task. Once participants pressed the SPACEBAR within the time limits, they were asked to decide whether a word presented onscreen was a real word of English or not. They had to use the F ('not a real word') and J ('real word') keys to make their decision. They had a maximum of 2000 milliseconds to press a key. If they failed to respond within this time, the trial would automatically end, participants would be asked to respond faster next time and the next trial would begin. Participants first went through two practice trials before the actual experiment began. They then saw filler and target items, which appeared in randomized order.

## Analysis

To analyze the data we used the R programming language (R Core Team, 2020) and R-Studio (RStudio Team, 2020). For data processing, visualization and analysis, we used the following packages: ggplot2 (Wickham, 2016), lme4 (Bates et al., 2007), Rmisc (Hope, 2013), MASS (Ripley et al., 2013), dplyr (Wickham et al., 2020), DoBy (Højsgaard, 2012), papaja (Aust and Barth, 2017), here (Müller, 2017), and afex (Singmann et al., 2020).

Prior to inferential analysis, we removed all participants who failed to accurately respond to the lexical decision task at least 70% of the time (across critical and filler trials). We also removed all critical trials for which participants gave a wrong answer. We then log-transformed the reaction times of the lexical decision task given that the residuals of a model using raw-reaction times were not normally distributed, and used log-milliseconds as our dependent variable.

We fitted a linear, mixed-effects model to the log-transformed data following the recommendations of Barr et al. (2013). The model included the four factors (METAPHOR TYPE, CONTEXT, FEATURE TYPE and ISI) and all possible interactions as fixed effects. Three factors (CONTEXT, FEATURE TYPE and ISI) had an ANOVA-style, sum-contrast coding, whereas the fourth factor METAPHOR TYPE was treatment-contrast coded, with the factor 'single metaphors' as the baseline. This allowed us to assess the three-way interaction between CONTEXT, FEATURE TYPE and ISI in the single metaphor case. This model was then re-fitted using the level 'extended metaphors' of the METAPHOR TYPE factor as the baseline. This second version of the model allowed us to examine the three-way interaction between CONTEXT, FEATURE TYPE and ISI for single metaphors, as well as the four-way interaction between METAPHOR TYPE, CONTEXT, FEATURE TYPE and ISI.

The random effects structure included random intercepts by items and by subjects. It also included random slopes for all factors and all possible interactions by items. The by-subjects random effects included slopes for all factors and all two- and three-way interactions.

## Predictions

Our predictions relate both to the comparison between single and extended metaphors, as well as to the replication of the original results of Rubio Fernández (2007). Rubio Fernández (2007) reported a loss of activation for superordinates (what we refer to as literal-related features) between 400 and 1,000 milliseconds relative to the activation levels of distinctive properties (referred to as metaphor-related features in the current study), which remained activated even after 1,000 milliseconds. Since we did not include a 400 millisecond level, we took the observed difference between 0 and 1,000 milliseconds in Rubio Fernández's experiment as the basis for the replication. For this reason, we considered that if the three-way interaction between CONTEXT, FEATURE TYPE and ISI was significant for the case of single metaphors, it would suggest that the activation levels of literal-related and metaphor-related features change as a function of time after processing single metaphors, in line with the original findings of Rubio Fernández (2007). If, on the other hand, this interaction is not significant, it would be at odds with the results of the original study. Our second prediction refers to the comparison between single and extended metaphors. Recall that we take the 'lingering of the literal' account proposed by Carston (2010) to predict qualitative differences in terms of the mechanisms involved in metaphor processing: Single metaphors are understood *via* the lexical modulation of the metaphoric vehicle, while during the comprehension of extended metaphors the literal meaning of the metaphor is maintained as a whole, with inferences projected later downstream in the form of an array of weak implicatures. In terms of the activation levels of literal-related features, the 'lingering of the literal' could translate to enhanced activation of these associated features given how both the literal meanings of the multiple related vehicles and features that are associated with them prime each other. This would mean that the three-way interaction between CONTEXT, FEATURE TYPE and ISI should not be significant for extended metaphors: the way in which activation levels of metaphor-related and literal-related features changes over time (relative to the unrelated baseline provided by the 'mismatch' conditions) should not be different from one another. This pattern should be accompanied by a significant four-way interaction between all four factors. This would suggest that while literal-related

features are suppressed with time when understanding single metaphors (supporting the lexical modulation of the meaning of the vehicle), these associated features would remain activated in the extended metaphor case, where the literal meaning would be metarepresented as a whole.

Alternatively, if we fail to find a significant four-way interaction and instead find similar three-way interactions between CONTEXT, FEATURE TYPE and ISI for both single and extended metaphors, it would suggest that the underlying mechanisms involved when processing single and extended metaphors are similar, contra Carston (2010).

## Results

The pattern of results is shown in Figure 2, while the summary of the model's output is shown in Tables 1, 2. Figure 2 shows the results in terms of 'priming time' for illustration purposes only, following the original reporting of results in Rubio Fernández (2007). This measure was calculated by subtracting the average values of response times in each 'match' from its corresponding 'mismatch' condition of the factor CONTEXT by items. By doing this, we obtained an estimate of the 'priming time' of each target word relative to the control baseline: Positive numbers represent a facilitatory effect (i.e., a positive priming effect on the lexical decision task of the target word), whereas negative numbers represent an inhibitory effect (i.e., a negative priming effect).

Table 1 shows the results for single metaphors (i.e., with 'single metaphors' as the baseline level for the factor METAPHOR TYPE). Here, we find a three-way interaction between CONTEXT, FEATURE TYPE and ISI ($t$-value = 3.3, $p < 0.005$). As can be seen in Figure 2, this interaction suggests that in the one-second difference in ISI, literal-related features are significantly reduced in activation relative to metaphor-related features (when comparing lexical decision times following the metaphor primes to lexical decision times of the same target words following unrelated controls). The results summarized in Table 2, which show the overall pattern for extended metaphors (i.e., with 'extended metaphors' as the baseline level for the factor METAPHOR TYPE), paint a different picture. Here, we failed to find a significant three-way interaction between CONTEXT, FEATURE TYPE and ISI ($t$-value = 1.1, $p = 0.26$). Crucially, there was a significant four-way interaction between all factors ($t$-value = 3, $p < 0.005$), suggesting that the pattern of activation of literal-related and metaphor-related features is different for extended relative to single metaphors. As Figure 2 suggests, it does not seem to be the case that literal-related features are suppressed with the change in ISI, in opposition to what seems to happen during processing of single metaphors.

## General discussion

In the current work, we set out to test the potential implications of Robyn Carston's 'lingering of the literal' account for the processing of extended metaphors. The account postulates a difference in processing strategies between extended and single metaphors. Processing single metaphors, according to the standard view within Relevance Theory, depends on the rapid construction of *ad hoc* categories. However, according to Carston (2010), relying on this mechanism might turn out to be overly strenuous for comprehenders when faced with an extended metaphor. This is because an extended metaphoric passage has multiple metaphoric vehicles that share the same literal conceptual domain. These multiple vehicles likely reinforce each other's literal meaning, leading comprehenders to maintain a representation of the literal meaning of the expression as a whole instead of relying on the lexical modulation of each vehicle individually. This account can explain why it has been consistently reported that there is a difference in processing between understanding metaphors preceded by other metaphors from the same conceptual domain relative to understanding the same metaphors presented in

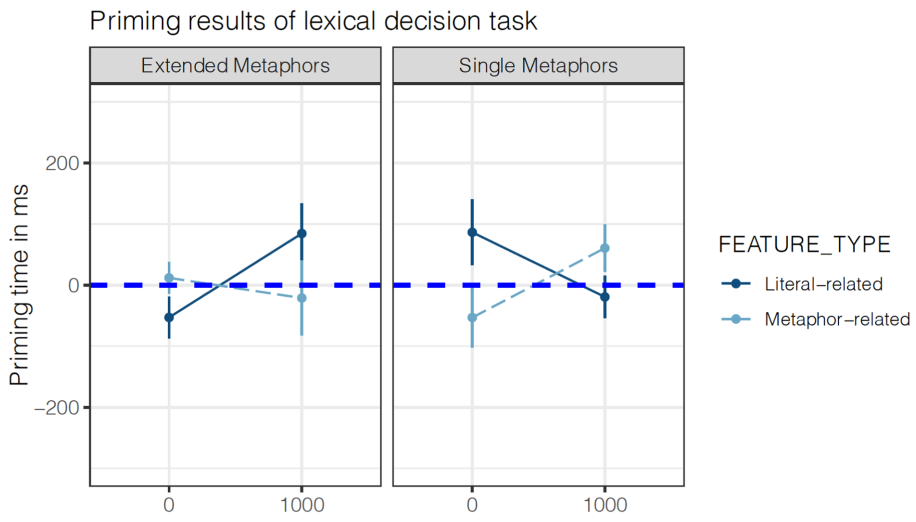



FIGURE 2
Activation pattern of literal-related and metaphor-related features. 'Priming time' was calculated by subtracting the 'match' level of the factor CONTEXT from the 'mismatch' level. This yielded the difference in milliseconds between processing the target word in the presence vs. absence of a related metaphor. Error bars show Standard Errors.

TABLE 1 Summary of regression model output with 'single metaphors' as baseline condition.

| Term | $\beta$ | 95% CI | $t$ | df | $p$ |
|---|---|---|---|---|---|
| ISI | 0.06 | [0.01, 0.11] | 2.42 | 12.82 | 0.031 |
| METAPHOR TYPE | −0.01 | [−0.05, 0.02] | −0.79 | 16.18 | 0.442 |
| FEATURE TYPE | 0.02 | [−0.05, 0.09] | 0.63 | 47.50 | 0.533 |
| CONTEXT | 0.03 | [−0.01, 0.07] | 1.63 | 17.02 | 0.122 |
| ISI*METAPHOR TYPE | −0.04 | [−0.10, 0.02] | −1.41 | 627.06 | 0.160 |
| ISI*FEATURE TYPE | 0.01 | [−0.07, 0.09] | 0.29 | 347.83 | 0.769 |
| METAPHOR TYPE*FEATURE TYPE | −0.06 | [−0.17, 0.05] | −1.07 | 369.94 | 0.286 |
| ISI*CONTEXT | 0.01 | [−0.07, 0.09] | 0.27 | 344.06 | 0.785 |
| METAPHOR TYPE*CONTEXT | −0.05 | [−0.11, 0.00] | −1.87 | 339.68 | 0.062 |
| FEATURE TYPE*CONTEXT | 0.01 | [−0.07, 0.09] | 0.28 | 15.98 | 0.787 |
| ISI*METAPHOR TYPE*FEATURE TYPE | −0.05 | [−0.17, 0.07] | −0.81 | 645.37 | 0.421 |
| ISI*METAPHOR TYPE*CONTEXT | −0.09 | [−0.22, 0.03] | −1.47 | 17.69 | 0.159 |
| ISI*FEATURE TYPE*CONTEXT | 0.26 | [0.11, 0.42] | 3.29 | 343.06 | 0.001 |
| METAPHOR TYPE*FEATURE TYPE*CONTEXT | 0.04 | [−0.06, 0.15] | 0.80 | 342.91 | 0.425 |

METAPHOR TYPE was treatment-contrast coded, all other factors were sum-contrast coded.

TABLE 2 Summary of regression model output with 'extended metaphors' as baseline condition.

| Term | $\beta$ | 95% CI | $t$ | df | $p$ |
|---|---|---|---|---|---|
| ISI | 0.02 | [−0.04, 0.08] | 0.69 | 16.61 | 0.499 |
| METAPHOR TYPE | 0.01 | [−0.02, 0.05] | 0.75 | 22.62 | 0.461 |
| FEATURE TYPE | −0.03 | [−0.10, 0.03] | −1.01 | 541.60 | 0.312 |
| CONTEXT | −0.02 | [−0.06, 0.02] | −0.95 | 19.47 | 0.352 |
| ISI*METAPHOR TYPE | 0.04 | [−0.02, 0.10] | 1.43 | 630.65 | 0.154 |
| ISI*FEATURE TYPE | −0.04 | [−0.12, 0.05] | −0.79 | 290.65 | 0.433 |
| METAPHOR TYPE*FEATURE TYPE | 0.05 | [−0.07, 0.17] | 0.76 | 64.33 | 0.449 |
| ISI*CONTEXT | −0.08 | [−0.16, 0.00] | −1.95 | 386.61 | 0.052 |
| METAPHOR TYPE*CONTEXT | 0.05 | [0.00, 0.11] | 1.82 | 368.06 | 0.070 |
| FEATURE TYPE*CONTEXT | 0.05 | [−0.03, 0.14] | 1.23 | 19.00 | 0.234 |
| ISI*METAPHOR TYPE*FEATURE TYPE | 0.05 | [−0.07, 0.16] | 0.76 | 646.20 | 0.449 |
| ISI*METAPHOR TYPE*CONTEXT | 0.09 | [−0.02, 0.20] | 1.55 | 376.52 | 0.121 |
| ISI*FEATURE TYPE*CONTEXT | −0.09 | [−0.25, 0.07] | −1.11 | 385.36 | 0.266 |
| METAPHOR TYPE*FEATURE TYPE*CONTEXT | −0.04 | [−0.15, 0.07] | −0.71 | 371.25 | 0.477 |
| ISI*METAPHOR TYPE*FEATURE TYPE*CONTEXT | 0.35 | [0.12, 0.57] | 3.04 | 379.25 | 0.003 |

METAPHOR TYPE was treatment-contrast coded, all other factors were sum-contrast coded

isolation (see also Gentner et al., 2001, for an alternative account). These empirical findings, however, can be explained *via* other accounts as well. Within Relevance Theory, for example, Wilson (2018) has claimed that exposure to extended metaphors brings about a 'lingering of linguistic form'. One possibility is that this involves a low-level facilitation effect that does not require the literal meaning of an expression to be meta-represented and does not bring about a qualitative difference in processing extended relative to single metaphors. In other words, it could be that comprehenders rely on the same mechanisms for understanding single and extended metaphors, and simple low-level priming that facilitates lexical access explains differences in comprehension time without requiring different mechanisms. Therefore, because existing empirical findings are compatible with multiple accounts, it is necessary to produce a stronger test of Carston's account. Specifically, one that can help determine whether the difference in processing single and extended metaphors is really caused by qualitative differences in the underlying comprehension mechanisms. The present work is a step in this direction. Based on the study by Rubio Fernández (2007), we set out to examine whether single and extended metaphors produce differences in the levels of activation of literal-related vs. metaphor-related features associated with the metaphor vehicle. We found that

during comprehension of single metaphors, metaphor-related features of the vehicle remain activated 1,000 milliseconds after metaphor comprehension while literal-related features show reduced activation, supporting the original findings of Rubio Fernández (2007). Critically, this was not the case for extended metaphors. Here, we found that literal-related features remain activated 1,000 milliseconds after the metaphor has been understood, on par with metaphor-related features. This finding supports Carston's 'lingering of the literal' account because it suggests that suppression of literal-related features is reduced or may not place when understanding extended metaphors. This mechanism has been claimed to be critically engaged during the comprehension of single metaphors by Gernsbacher et al. (2001) and Rubio Fernández (2007), and the fact that suppression of literal-related features might not be at play for extended metaphors suggests a more prominent role of literal meaning in the interpretation of this type of metaphor. This may point to qualitative differences in the underlying comprehension mechanisms–e.g., in the form of different "modes of metaphor processing" as suggested by Carston (2010).

The current findings come with some caveats and are to be interpreted with caution. First of all, adapting Rubio Fernández's paradigm to a web-based task led us to change the cross-modal priming design and present both prime and target visually. This has the limitation that we cannot be certain of the exact moment during processing in which the target word is read relative to the metaphoric prime. Though we attempted to account for this fact by setting a maximum amount of time for participants to read the metaphoric prime (4,000 milliseconds), it remains less than ideal. For a better reduction of noise, it would be necessary to run the experiment in the lab as a cross-modal priming task. A further constraint of the web-based paradigm is naturally also the reduced number of critical items used. Expanding this number would allow for better generalizability. Future research should also investigate potential differences in suppression as a function of the number of metaphoric vehicles that comprehenders are faced with. This would help answer the question of the point in time in which we would expect comprehenders to 'switch' from one processing mode to another (assuming that extended metaphors are actually processed differently than single ones).

Another important caveat concerns the linking assumptions used in the current experiment. In this work, we laid out an interpretation, in processing terms, of both the 'lingering of the literal' and 'lingering of linguistic form' accounts. However, these accounts are underspecified from a processing point of view, and are in theory compatible with various different empirical predictions. For example, it could be that the 'lingering of linguistic form' also predicts a suspension of the mechanism of suppression for literal-related features, if this view were to be interpreted differently than we have in the current work. Further work is therefore needed from both theoretical and experimental perspectives in order to thoroughly explicate the links between theory and processing and to solve the 'puzzle' of extended metaphors.

## Conclusion

Since the work by Carston (2010), extended metaphors have represented an interesting battle ground in the development of the relevance-theoretic view on metaphor comprehension. In the current work, we provide a new type of empirical evidence in favor of the 'lingering of the literal' account. Our experiment suggests that comprehenders do not suppress literal-related features when understanding extended metaphors (but they do so when understanding single metaphors). This in turn supports the idea that understanding extended metaphors involves maintaining a representation of the literal meaning of the entire metaphoric expression.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://osf.io/eayj7/.

## Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

CR: conceptualization, writing—original draft preparation, and writing—review and editing. IF: writing—review and editing and supervision. All authors contributed to the article and approved the submitted version.

## Funding

This paper is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 853211 (ERC Starting Grant 2019), awarded to IF.

## 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.

## Publisher's note

# References


Aust, F., and Barth, M.. (2017). *Papaja: Prepare Reproducible APA Journal Articles with R Markdown*. R Package Version No. 0.1.0.9997.

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., Sarkar, D., Bates, M. D., and Matrix, L.. (2007). *The Lme4 Package*. R Package Version, No. 2.74.

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., and Chatterjee, A. (2012). From novel to familiar: tuning the brain for metaphors. *NeuroImage* 59, 3212–3221. doi: 10.1016/j.neuroimage.2011.11.079

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. New York: John Wiley & Sons.

Carston, R.. (2010). *XIII-Metaphor: Ad hoc Concepts, Literal Meaning and Mental Images*. Proceedings of the Aristotelian Society (Hardback), 110, pp. 295–321.

Futrell, R. (2012). *Processing Effects of the Expectation of Informativity*. Stanford University. Master's Thesis.

Gentner, D., and Bowdle, B. (2008). "Metaphor as structure-mapping" in *The Cambridge Handbook of Metaphor and Thought*. ed. R. Gibbs (New York: Cambridge University Press), 109–128.

Gentner, D., Bowdle, B. F., Wolff, P., and Boronat, C. (2001). "Metaphor is like analogy" in *The Analogical Mind: Perspectives from Cognitive Science*. eds. D. Centner, K. J. Holyoak and B. N. Kokinov (Cambridge, MA: The MIT Press), 199–253.

Gernsbacher, M. A., and Faust, M. E. (1991). The mechanism of suppression: a component of general comprehension skill. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 245–262. PMID: 1827830

Gernsbacher, M. A., Keysar, B., Robertson, R. R. W., and Werner, N. K. (2001). The role of suppression and enhancement in understanding metaphors. *J. Mem. Lang.* 45, 433–450. doi: 10.1006/jmla.2000.2782

Glucksberg, S. (2001). *Understanding Figurative Language*. Oxford: Oxford University Press.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends Cogn. Sci.* 7, 92–96. doi: 10.1016/S1364-6613(02)00040-2

Glucksberg, S. (2008). "How metaphors create categories–quickly" in *The Cambridge Handbook of Metaphor and Thought*. ed. R. Gibbs (New York: Cambridge University Press), 67–83.

Højsgaard, S. (2012). *The doBy Package*. R Package Version, No. 4.

Holyoak, K. J., and Stamenković, D. (2018). Metaphor comprehension: a critical review of theories and evidence. *Psychol. Bull.* 144, 641–671. doi: 10.1037/bul0000145

Hope, R. M. (2013). *Rmisc: Ryan Miscellaneous*. R Package Version, No. 1.

Keysar, B., Shen, Y., Glucksberg, S., and Horton, W. S. (2000). Conventional language: how metaphorical is it? *J. Mem. Lang.* 43, 576–593. doi: 10.1006/jmla.2000.2711

McGlone, M. S., and Manfredi, D. A. (2001). Topicvehicle interaction in metaphor comprehension. *Mem. Cogn.* 29, 1209–1219. doi: 10.3758/BF03206390

Müller, K. (2017). *Here: A Simpler way to Find your Files (Version 0.1)*. Available at: https://CRAN.R-project.org/package=here

Pouscoulous, N., and Dulcinati, G. (2019). "Metaphor" in *Oxford Handbook of Experimental Semantics and Pragmatics*. eds. C. Cummins and N. Katsos (Oxford: Oxford University Press)

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., et al. (2013). Package "mass". *Cran R* 538, 113–120.

Ronderos, C. R., Guerra, E., and Knoeferle, P. (2021). The role of literal features during processing of novel verbal metaphors. *Front. Psychol.* 11:3899. doi: 10.3389/fpsyg.2020.556624

Ronderos, C. R., Guerra, E., and Knoeferle, P. (2022). When sequence matters: the processing of contextually biased German verb–object metaphors. *Lang. Cogn.* 15, 1–28. doi: 10.1017/langcog.2022.22

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC.

Rubio Fernández, P. (2007). Suppression in metaphor interpretation: differences between meaning selection and meaning construction. *J. Semant.* 24, 345–371. doi: 10.1093/jos/ffm006

Rubio-Fernández, P., Cummins, C., and Tian, Y. (2016). Are single and extended metaphors processed differently? A test of two relevance-theoretic accounts. *J. Pragmat.* 94, 15–28. doi: 10.1016/j.pragma.2016.01.005

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., et al. (2020) *Package Afex*. Available at: https://cran.r-project.org/web/packages/afex/afex.pdf

Sperber, D., and Wilson, D. (2008). "A deflationary account of metaphors" in *The Cambridge Handbook of Metaphor and Thought*. ed. R. Gibbs (New York: Cambridge University Press), 84–105.

Thibodeau, P., and Durgin, F. (2008). Productive figurative communication: conventional metaphors facilitate the comprehension of related novel metaphors. *J. Mem. Lang.* 58, 521–540. doi: 10.1016/j.jml.2007.05.001

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham, H., François, R., Henry, L., and Müller, K. (2020). *Dplyr: A Grammar of Data Manipulation*. Available at: https://CRAN.R-project.org/package=dplyr

Wilson, D. (2018). "Relevance theory and literary interpretation" in *Reading Beyond the Code: Literature and Relevance Theory*. eds. T. Cave and D. Wilson (Oxford, UK: Oxford University Press)

Wilson, D., and Carston, R. (2006). Metaphor, relevance and the emergent property issue. *Mind Lang.* 21, 404–433. doi: 10.1111/j.1468-0017.2006.00284.x

Wilson, D., and Carston, R. (2007). "A unitary approach to lexical pragmatics: relevance, inference and ad hoc concepts" in *Pragmatics*. ed. N. Burton-Roberts (London: Palgrave Macmillan UK), 230–259.

Wilson, D., and Sperber, D. (2012). *Meaning and Relevance*. Cambridge: Cambridge University Press.

Zehr, J., and Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 11 April 2023
DOI 10.3389/fpsyg.2023.1116842



OPEN ACCESS

EDITED BY
Tim Wharton,
University of Brighton,
United Kingdom

REVIEWED BY
Kaja Borthen,
Norwegian University of Science and Technology,
Norway
Ryoko Sasamoto,
Dublin City University,
Ireland

*CORRESPONDENCE
Line Sjøtun Helganger
✉ line.s.helganger@usn.no

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

RECEIVED 05 December 2022
ACCEPTED 27 February 2023
PUBLISHED 11 April 2023

CITATION
Helganger LS and Falkum IL (2023) Intonational production as a window into children's early pragmatic competence: The case of the Norwegian polarity focus and two *jo* particles.
*Front. Psychol.* 14:1116842.
doi: 10.3389/fpsyg.2023.1116842



# Intonational production as a window into children's early pragmatic competence: The case of the Norwegian polarity focus and two *jo* particles

Line Sjøtun Helganger[1]* and Ingrid Lossius Falkum[2,3]

[1]Department of Languages and Literature Studies, University of South-Eastern Norway, Kongsberg, Norway, [2]Department of Philosophy, Classics, History of Arts and Ideas, University of Oslo, Oslo, Norway, [3]Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway

The use of the Norwegian intonation pattern Polarity Focus highlights the polarity of a contextually given thought and enables the speaker to signal whether she believes it to be a true or false description of some state of affairs. In this study, we investigate whether preschool children can produce this intonation pattern and what their productions reveal about the development of their early pragmatic abilities. We also explore their use of Polarity Focus in combination with two particles encoded by the linguistic form *jo*: a sentence-initial response particle, and a sentence-internal pragmatic particle. We used a semi-structured elicitation task consisting of four test conditions of increasing complexity to shed light on the developmental trajectory of the mastery of Polarity Focus. Our results show that already from the age of 2 children are proficient users of this intonation pattern, which occurs in three out of four conditions for this age group. As expected, only 4- and 5-year-olds produced Polarity Focus in the most complex test condition that required the attribution of a false belief. We further found production of sentence-initial response particle *jo* by all age groups, both in combination with Polarity Focus and alone. Production of the sentence-internal pragmatic particle *jo*, felicitously co-occurring with Polarity Focus, emerges around age 3. This study presents the first experimental evidence of Norwegian children's mastery of intonation as a communicative device in language production and their use of the two *jo* particles. We show how intonational production can be used as a window into children's early pragmatic competence: The mastery of the production of Polarity Focus can be seen as an early linguistic manifestation of the cognitive abilities for the attribution of thoughts and epistemic vigilance towards propositional content.

KEYWORDS

**pragmatic development, intonation, polarity focus, response particles, pragmatic particles, metarepresentation, epistemic vigilance, relevance theory**

## 1. Introduction

Imagine that you are talking to your three-year-old who is trying to explain that he sent a letter to his buddy at the kindergarten earlier that day. To make sure that you have understood what he has just told you, you say: *Så han var ikke i barnehagen?* ('So, he was not at kindergarten?'). The child replies as in (1) below:[1]

(1) Å (3;10): ((han($^{1}$VAR-i$_{AP}$)$_{IP}$)($^{2}$barnehagen$_{AP}$)$_{IU}$)
L* H$^{-}$ H*L (H) L%
he WAS-in kindergarten$_{DEF}$
'He was at kindergarten (despite what you are saying).'

In (1), the child provides the information that his friend was present at the kindergarten that day. However, the child also signals that there is an opposition between what you seem to think (that his friend was not present at the kindergarten) and the actual state of affairs (that his friend was present at the kindergarten). The child denies the truth of the proposition expressed by your utterance, and he does so by using a particular intonation pattern, the so-called Polarity Focus (PF; e.g., Fretheim, 2002), characterized by a focal accentuation of a polarity carrier (in (1), the finite verb *var* ('was')) followed by an additional accentuation later in the utterance (in (1), *barnehagen* ('the kindergarten')). What is particular about such PF utterances compared to utterances realized without PF, is that the use of PF signals that the only new information in the utterance is the truth value of the proposition expressed by the speaker. This makes PF a valuable tool for a speaker who wants to convince her interlocutor of whether a contextually given thought is a true or false description of some state of affairs.

For a speaker to be able to signal a denial of her interlocutor's thought, it must be manifest to her that her interlocutor believes this thought, that is, she must be capable of mentally representing that her interlocutor believes this thought and accepting this as true or probably true (Sperber and Wilson, 1986/1995, p. 39; Wilson, 2012). The use of PF therefore requires the abilities to metarepresent and attribute thoughts. Also, since the production of PF arguably requires the speaker to be attentive to, evaluate, and express an attitude (of endorsement or denial) toward the truth-conditional content of an attributed thought, it is a higher-level metarepresentational ability that is required (*cf.* Wilson, 2012), in addition to a capacity for 'epistemic vigilance' toward utterance content (*cf.* Sperber et al., 2010). The example in (1) above (taken from the first author's diary notes of her son) suggests that PF may occur early in children's language production. In this study, we ask whether preschool children tend to produce this intonation pattern, and, if so, what their production of PF utterances can reveal about the development of their early pragmatic abilities.

Instead of (1), the child could also have had responded as in (2a), (2b) or (2c) below to communicate the same (or similar) content:

(2) a. A: *Så han var ikke i barnehagen?*
'So, he was not at the kindergarten?'
B1: (((JO$_{AP}$)$_{IP}$)$_{IU}$)
Yes$_{RESP.PART}$
'Yes (contrary to what you are saying, he was at the kindergarten).'
b. B2: (((JO$_{AP}$)$_{IP}$)$_{IU}$), ((han($^{1}$VAR-i$_{AP}$)$_{IP}$)($^{1}$barnehagen$_{AP}$)$_{IP}$)$_{IU}$)
Yes$_{RESP.PART}$ he WAS-in kindergarten$_{DEF}$
'Yes, he was at the kindergarten (despite what you seem to think).'
c. B3: ((han($^{1}$VAR-jo-i$_{AP}$)$_{IP}$)($^{1}$barnehagen$_{AP}$)$_{IU}$)
he WAS-$_{PART}$-in kindergarten$_{DEF}$
'You know, he was at kindergarten (despite what you seem to think).'

In addition to investigating PF production, this study explores children's use of two particles, both orthographically expressed as *jo*, which often co-occur with PF and share some of its pragmatic features: they both enable the speaker to signal her attitude toward the truth value of the proposition expressed. *Jo* can appear either as a sentence-initial response particle (Fretheim, 2017), such as in (2ab) above, or as a sentence-internal pragmatic particle (Berthelin and Borthen, 2019), as in (2c).[2]

According to Fretheim (2017), the sentence-initial response particle *jo* is used to contradict a communicated negation by affirming the embedded positive proposition. Thus, both PF and the response particle *jo* require that the speaker metarepresents a contextually manifest thought and they both relate to the truth of the proposition expressed.[3] The response particle *jo* can be used as a single word response, such as in (2a) above, or it can precede a PF utterance, such as in (2b).

As to the sentence-internal pragmatic particle *jo*, Berthelin and Borthen (2019) suggest that its semantic contribution is to signal that the proposition expressed should be interpreted as mutually manifest to the speaker and hearer, and to signal that the utterance can be taken as a premise for deriving a contextual implication. Sentence-internal *jo* therefore naturally accompanies PF in (2c), providing the hearer

1 The utterance is transcribed using the parentheses notation convention of the Trondheim model (e.g., Nilsen, 1992; Fretheim, 2002), a framework developed for analysis of Norwegian intonation. The line below the transcription contains an annotation of the utterance's realized tones, and the two last lines are the English translation of the utterance. Capital letters indicate a focal accentuation characterized by an initial realization of a lexical word tone marked by a parenthesis-initial $^{1}$ for Accent 1 (L*) or $^{2}$ for Accent 2 (H*L), followed by a tonal rise to an extra high tone (H$^{-}$) marked by a parenthesis-final $_{AP}$ followed by an $_{IP}$. Norwegian utterances must have at least one, and never more than two, focal accentuations. This tonal movement, starting with the onset of a word tone and then rising from the L(ow) tone to an extra high – focal – tone (H$^{-}$), is the most important information structural feature of (East) Norwegian intonation. Focal accentuations signal which part of the utterance is to be interpreted as new information, and which part constitutes information that provides the hearer with relevant context to interpret the new information against (e.g., Fretheim, 2002). Non-focal accentuation is also realized with an initial lexical word tone ($^{1}$ or $^{2}$) and is followed by a tonal rise to a high, but not extra high, tone ((H)), marked in the transcription by a parenthesis-final $_{AP}$. Unaccented segments have no lexical word tone and are separated from accented words either by an opening parenthesis to the right or by a dash (–) to the left in the transcriptions.

2 Note that the response particle (i.e., the sentence-initial) *jo*, and the pragmatic particle (i.e., the sentence-internal) *jo* have different etymologies (*cf.* Berthelin and Borthen, 2019, pp. 4, 27fn2).

3 Note, however, that whereas the relevance of *jo* is limited to preceding contexts containing a negative surface structure, PF is possible as a response to both negative and positive surface structures.

with an additional cue to the speaker's intention: that the hearer should accept as true that the speaker's friend was indeed present at the kindergarten. Because of the overlapping pragmatic features between PF and the *jo* particles described above, more careful analyses of these early *jo* productions may serve to cast light on the development of metarepresentational abilities in children.

In (1) above, PF was used to signal the speaker's denial of the truth of the metarepresented thought. However, PF can also be used to affirm the truth of a thought, as in (3) below where A concludes after hearing that B's fried was present at kindergarten when B had sent his letter:

(3) A: *Da fikk han brevet med en gang.*
'Then he received the letter right away.'
B: $(((^{1}\text{han}_{AP})(^{1}\text{FIKK}_{AP})_{IP})(^{1}\text{brevet-med}_{AP})(^{1}\text{en}_{AP})(^{1}\text{gang}_{AP})_{IU})$
L*(H) L* H– L* (H) L*(H) L*(H) L%
He RECEIVED $\text{letter}_{DEF}$-with one time
'He DID receive the letter right away (just as you seem to think)!'

The PF in B's utterance is realized as a focal accentuation of the finite verb *fikk* ('received') followed by three additional non-focal accentuations. Using PF, B metarepresents and affirms the truth of the proposition explicitly expressed by A in (3).

PF can also be used to affirm or deny a thought that is not explicitly mentioned but attributed to someone (or to oneself). Consider the conversation in (4) below:

(4) A: *Jeg kommer meg ikke til butikken!*
I come me not to $\text{grocery store}_{DEF}$
'I can't get to the grocery store!'
B (who knows A has an electric car):
$(((^{1}\text{bilen}_{AP})(^{1}\text{ER}_{AP})_{IP})((^{2}\text{LADET}_{AP})_{IP})_{IU})$
L*(H) L*H– H*L H-H%
$\text{car}_{DEF}$ IS CHARGED
'The car is charged (despite what you seem to think)'.

Here B's utterance is realized with PF, involving a focal accentuation of the finite verb *er* ('is') followed by another (focal) accentuation of the infinite verb form *ladet* ('charged'). A expresses that he cannot get to the grocery store, and by responding with a PF utterance, B communicates that she denies a (false) belief that she attributes to A: that he cannot use his (electric) car to drive to the grocery store because it is discharged.

The analysis of the use and function of PF has an affinity with the Relevance Theoretic notion of 'echoic utterances'. Using an echoic utterance, the speaker metarepresents and attributes a thought or utterance with a similar content to someone else (or to the speaker herself), and at the same time conveys the speaker's attitude to this thought or utterance. It is this signaling of the speaker's attitude to the attributed thought, which is characteristic of echoic utterances, and by which they achieve relevance (Wilson, 2012, p. 249). Using an echoic utterance is one linguistic tool available to the speaker if she intends to modify her interlocutor's epistemic state. Utterances carrying PF serve a similar communicative function in that they involve the affirmation or denial of an attributed proposition as a description of some state of affairs, and could in this way be said to involve an echoic element.

An overarching aim of this study is to gain a deeper understanding of intonation as part of a broader pragmatic competence. According to Relevance Theory (Sperber and Wilson, 1986/1995), intonation serves an important pragmatic function by contributing to an utterance's relevance (where 'relevance' is understood as a trade-off between so-called cognitive effects and processing effort). On this view, intonation functions procedurally as a guide to the speaker's intended interpretation:[4] by signaling an utterance's information structure it contributes to making some contextual implications more salient than others (Fretheim, 2002; Wilson and Wharton, 2006; Wilson, 2011; Scott, 2021).

The early emergence of prosodic competence in first language acquisition (for an overview, see Kehoe (2013)) and the pragmatic nature of intonation provide us with an opportunity to use intonation as a window into children's early pragmatic competence. Children's acquisition of intonation in the period prior to five years of age is still a quite unexplored field of research (Peppé and Wells, 2014). Furthermore, although there are studies on the role of (intonational) focus in pragmatic reasoning (e.g., Tomlinson et al., 2017; Gotzner and Spalek, 2019), there have been few attempts to combine suprasegmental phonology with cognitive pragmatic theory in the study of language acquisition (Wharton, 2012, 2020). Thus, the question of how children's ability to master intonation as a communicative device develops remains largely unresolved. Analyses of children's production of intonation utterances can provide us with a deeper understanding of what this ability amounts to. In addition, this study presents the first attempt of accounting for Norwegian speaking children's use of intonation as a communicative device in language production.

## 1.1. Previous research

While some developmental studies have investigated children's ability to produce intonational focus (e.g., Wieman, 1976; Wells et al., 2004; Romøren and Chen, 2021), we know of no previous studies that have specifically investigated children's production of PF. However, the literature discusses a similar intonation pattern, often referred to as 'Verum Focus' (e.g., *The house ISn't on fire*; Gussenhoven, 1983, p. 406). In an elicitation experiment, Turco et al. (2014) showed adult participants pictures of different situations (e.g., a man washing a car). Participants then heard prerecorded utterances where the depicted situations were negated (e.g., *The man is NOT washing the car*). Together, the visual and audio stimuli served as a context for eliciting Verum Focus utterances, where the truth value of the negation provided in the audio stimuli was to be corrected (e.g., *The man IS washing the car*). Results showed that that German adult speakers produced Verum Focus in more than 70% of the cases in these 'polarity correction' contexts.[5]

In an adaptation of Turco et al.' (2014) study, Dimroth et al. (2018) investigated German four-to six-year-olds' production of this intonation pattern but found only a small number of occurrences (5 out of 175 trials). In their adult control group, Verum Focus occurred in 53 out of 99 trials. However, despite several similarities between Verum Focus and PF, the two notions are not equivalent: Verum Focus is used in a broader sense, also including contexts where the polarity of a metarepresented proposition is not really at question, such as the 'polarity contrast' context of Dimroth et al. (2018). In this context, a confederate describes a picture only visible to him, using a negative utterance (e.g., *In my*

4 Cf. Blakemore (1987, 2002) and Wilson (2011, 2016) for more on the distinction between 'conceptual meaning' and 'procedural meaning'.

5 See also Turco et al. (2013) for a study comparing German and French realizations of Verum Focus accent in adult speakers.

*picture the child is not eating the candies*) and the participant's task is to respond by describing her own picture. This picture shows the affirmative version of the confederate's picture (e.g., a child eating candies), only accessible to the participant herself, leading to responses such as *On mine the child HAS eaten the candies* (Dimroth et al., 2018, p. 276). In this context, the accentuation of the finite verb does not highlight the polarity of any proposition; the issue is not whether or not it is true that the child in the participant's picture eats candies. Rather the participant's response highlights the difference (contrast) between the motive in the participant's and the confederate's pictures.

As to the Norwegian *jo* particles, they have not previously been studied in a developmental context. Noveck et al. (2021) investigated children's production of the French equivalent to the Norwegian sentence-initial response particle *jo*, the response particle *si*. They describe the response particle as "a pragmatically rich response that addresses the questioner's epistemic state" (*ibid.*, p. 4). *Si* can be used to respond affirmatively when a negative question at the surface structure turns out to be a false negative one, implicitly signaling the questioner's positive belief (e.g., *It is not in the white box?*). In Noveck and colleagues' study, participants answered a question of whether a candy was in a box or not. Each trial started with a puppet declaring his prior belief about the candy's whereabouts before the participant inspected the box. Then the puppet asked either an affirmative question (e.g., *It is in the white box?*) or a negative question (e.g., *It is not in the white box?*). Crucially, in the *si*-eliciting condition the puppet asked a negative question, but the box contained the candy.

The results showed that six-year-olds are adult-like in their uses of *si* but four-year-olds are not. Although the four-year-olds showed adult-like accuracy rates (where accuracy was understood in terms of pragmatically felicitous responses), answering *Oui*, *Non* and *Si* to the puppet's question just as correctly as the six-year-olds and adults did, they were strikingly faster than the six-year-olds and adults in responding *si* in the context of a false negative question. According to Noveck et al. (2021) this accurate, but unexpectedly fast response indicates that four-year-olds rely on a minimal semantic representation of *si* when answering the question (in rejecting the content of the false-negative question), but do not yet fully appreciate its pragmatic complexity which involves "[modifying] the questioner's epistemic state so that it aligns better with the answerer's" (*ibid.*, p. 22).

If Noveck et al. (2021) are right in their analysis of the four-year-olds' pragmatic immaturity–and assuming that the response particles *si*, *jo* and PF broadly serve the same pragmatic function of metarepresenting a contextually manifest thought and expressing an attitude toward the truth-conditional content of this thought–we should not expect four-year-olds, and certainly not children younger than four years of age, to produce PF.

However, we are not entirely convinced by the conclusion Noveck et al. (2021) draw regarding four-year-olds' pragmatically limited use of *si*. From the developmental literature, we know that already around the age of two, children have a capacity for metarepresentation (Leslie, 1987), they can reject false and accept true statements (Lyon et al., 2013), and they can spontaneously contradict and correct assertions that they believe to be false (Pea, 1982). Furthermore, already from around 14 months of age, children's perspective-taking abilities include the understanding that attitudes of others to objects of joint attention may differ from their own (O'Madagain and Tomasello, 2021). It seems puzzling to us that they would not also make use of these abilities when producing *si* in appropriate contexts. Arguably, the main informative intention of a speaker who uses this particle is to convey her denial of a metarepresented thought (why else would she use it?). Furthermore, it is likely that her goal in conveying this is to modify her interlocutor's epistemic state. This would seem to involve an understanding that goes beyond accessing the minimal semantic representation (i.e., the mere rejection of a negative surface structure), and which includes beginning mastery of the pragmatic processes involved in the mature use of the utterance to affirm the questioner's positive belief. It is this pragmatically rich understanding that seems to be involved in the use of utterances containing the response particle *jo*, PF, or a combination of the two, such as in (2b) above.

## 1.2. Hypotheses

We hypothesize that preschool children should be able to produce PF in appropriate contexts. This hypothesis is based partly on anecdotal observations of children's early PF productions from diary notes and private recordings, and partly on what we know about the early development of some of the prerequisite abilities for use of PF (*cf.* Pea, 1982; Leslie, 1987; Lyon et al., 2013; O'Madagain and Tomasello, 2021), as well as children's pragmatic sophistication in related domains such as the ability to draw scalar implicatures (Pouscoulous et al., 2007), to grasp presuppositions (Berger and Höhle, 2012), and to appropriately use referring expressions (Matthews et al., 2006). However, given that the ability to linguistically express an understanding of false beliefs appears around children's fourth birthday (Wellman et al., 2001), we would only expect children aged four years and older to produce PF in the most complex context where they have to infer and attribute a false (or ignorant) belief to their interlocutor (*cf.* (4) above). The study's hypotheses are preregistered in OSF: https://osf.io/3asu5/.

In the examples in (1)–(3) above, the proposition echoed is explicitly expressed by the interlocutor prior to the speaker's PF utterance and is therefore easily accessible. We hypothesize that the use of PF in such contexts is acquired earlier than in contexts where the proposition echoed must be inferred (such as in (4) above).

Findings from the developmental literature suggest that the presence of negation increases the complexity of utterances (Just and Carpenter, 1971; Clark and Chase, 1972). We hypothesize that use of PF to affirm a positive proposition is acquired earlier than the denial of a positive proposition, followed by the ability to deny a prior negative belief.

We expect the earliest starting point of PF production to be around two years of age, by the age typically developing children have usually started to produce word combinations (Kristoffersen and Simonsen, 2012). This hypothesis is based on the intonational criteria for PF production: An utterance realized with two accentuations must consist of at least two words (Fretheim and Nilsen, 1993). Furthermore, Lyon et al. (2013) have shown that children are able to accept and reject true and false statements before their second birthday, suggesting a developing ability for epistemic vigilance toward utterance content (Sperber et al., 2010).

# 2. Method

## 2.1. Participants

This study includes 92 children within the age range of 2;2 to 5;9 years, divided into four groups: two-year-olds ($n = 20$), three-year-olds ($n = 20$), four-year-olds ($n = 31$), and five-year-olds ($n = 21$). Seven additional participants were omitted from the analyses because they produced no

comprehensible multiword utterances during the recording sessions ($n=6$) or failed to concentrate on the experimental tasks ($n=1$). The participants had South-East Norwegian as their first language[6] and were recruited through kindergartens in the South-Eastern region of Norway. Prior to data collection the study received ethical approval from NSD–Norwegian Center for Research Data (project number 60923) and written parental consent was obtained. Participants were tested individually in a quiet room in the kindergarten or in their private home. To capture the intonational production of the children and as much of the context as possible, the participants were video recorded using a Sony video camera recorder HDR-CX410 with a 5.1ch surround microphone. Each session lasted for approximately 10 min.

## 2.2. Procedure and materials

Our semi-structured design involves an elicitation task combined with intermediate sections of spontaneous speech. The sections of spontaneous speech are included to make the experimental setting as similar as possible to a natural conversation. The initial unstructured conversation is especially important for establishing a relation between the participant and the experimenter, and for the participant to get acquainted with the handpuppet used in the elicitation task. It also serves to establish the relevant context for the elicitation task that follows.

First, an experimenter and a handpuppet show the participant some of the handpuppet's toys (three rubber ducks and a little ball) during an unstructured conversation, where the handpuppet demonstrates that he is a bit forgetful. The experimenter, the handpuppet and the participant play with and talk about the toys, commenting on how they look, what they can be used for, and so on. The handpuppet explicitly states that he loves playing with his rubber ducks. In the remainder of the unstructured sections, the participant, handpuppet and experimenter talk about topics related to the structured elicitation task they go through.

Second, participants are presented with the structured PF elicitation task. Inspired by the 'polarity correction' context of Turco et al. (2014), we used still-life pictures as visual stimuli in three of the conditions. From Noveck et al. (2021) we adapted the procedure whereby a puppet initially explicitly states his (positive or negative) prior belief in the form of a declaration regarding some state of affairs (e.g., *I believe that the boy is eating strawberries*) before the visual stimuli is presented. Depending on the condition, the prior belief is either a match or a mismatch as a description of the picture's motive. The crucial task for the participant is to produce a target utterance in response to the puppet's utterance about the motive in the picture. Before the picture is presented the puppet hides so he cannot see, making it more likely that the participant will produce an utterance. If the participant does not produce any utterance, the puppet, still hiding, will ask an elicitation question to prompt the child to produce an answer (e.g., *Does the boy eat strawberries?*). Participants are not given any kind of instructions for what or how to respond, it is their spontaneous production that is of interest.

In one of the conditions, instead of expressing a belief about the content of a picture the handpuppet expresses a desire (*I wish I had something to play with while taking a bath*), suggesting that he does not remember his rubber ducks (i.e., the ones they had played with in the initial conversation). Here the production of PF is only relevant as a response if the participant has drawn the necessary inferences about a (false) belief of the handpuppet (i.e., that he does not have rubber ducks).

To familiarize the children with the procedure of responding to the puppet's prior beliefs about the pictures, we included a familiarization trial. The four test conditions are thought to be of increasing complexity,[7] starting with the "Positive-Affirmation" condition as the simplest one where the child only has to affirm a positive proposition. Next is the "Positive-Denial" condition where a positive proposition must be denied. Third is the "Negative-Denial" condition which involves the contradiction of a negated proposition. Fourth, and with highest complexity, is the "Inferred Belief-Denial" condition where an inferred (negative) belief must be contradicted. This increasing complexity enables the study of a potential developmental trajectory of the mastery of PF. In addition we included a Control condition, where use of PF is not relevant because the context provides no proposition of which to highlight the polarity. All participants were tested in all conditions and each participant saw a total of five test items.

The conditions were pseudo-randomized. The Inferred Belief-Denial condition was set as the third trial across participants. We did this to ensure that (i) it did not occur too soon after the visual stimuli had been presented, (ii) that retrieving the visual stimuli from memory was not too effortful, and (iii) the memory demands were the same for all children. The remainder of the conditions were randomized. See (5)–(10) below for an overview of the study's familiarization trial and conditions:

(5) **Familiarization trial**
Introduction by experimenter: *Now you will see a picture.*
Prior belief (opinion) by puppet: *I love to watch pictures!*
Visual stimuli: A sleeping dog.
Elicitation question: *What do you see in the picture?*

(6) **Positive-Affirmation condition (Pos-Aff)**
Introduction by experimenter: *Next is a picture of a boy.*
Prior POSITIVE belief: *I believe that the boy is eating strawberries.*
Visual stimuli: Match–a boy eating strawberries.
Elicitation question: *Does the boy eat strawberries?*
Potential PF response: $(([^1\text{gutten}_{AP})([^1\text{SPISER}_{AP})_{IP})([^1\text{jordbær}_{AP})_{IU})$
$\text{boy}_{DEF}$ EATs strawberries
'The boy DOES eat strawberries.'

(7) **Positive-Denial condition (Pos-Den)**
Introduction by experimenter: *Next is a picture of a girl.*
Prior POSITIVE belief: *I believe that the girl is throwing a ball.*
Visual stimuli: Mismatch–a girl lying in the grass (witout any ball).
Elicitation question: *Does the girl throw a ball?*
Potential PF response: $(([^2\text{jenta}_{AP})([^2\text{KASTER-ikke}_{AP})_{IP})([^1\text{ball}_{AP})_{IU})$
$\text{girl}_{DEF}$ THROWS-not ball
'The girl does NOT throw (a) ball.'

(8) **Negative-Denial condition (Neg-Den)**
Introduction by experimenter: *Next is a picture of a boy.*
Prior NEGATIVE belief: *I believe that the boy is not reading a book.*
Visual stimuli: Mismatch – a boy reading a book.

6 South-East Norwegian is a dialect spoken in the South-Eastern region of Norway.

7 Noveck et al. (2021, p. 22) point out that in classic tasks where participants are given statements (e.g., *A robin is a bird*) and the options True versus False, reaction times typically increase in correlation with increasing number of negations (in the answer or the question).

Elicitation question: *Does the boy not read a book?*
Potential PF response: ((($^{1}$gutten$_{AP}$)($^{1}$LESER$_{AP}$)$_{IP}$)($^{1}$bok$_{AP}$)$_{IU}$)
boy$_{DEF}$ READS book
'The boy DOES read (a) book.'

(9) **Inferred belief-Denial condition (Inf-Bel)**
Verbal stimuli: Puppet: *I wish I had something to play with while taking a bath!*
(Visual stimuli: The rubber ducks used initially in the unstructured conversation)
Potential PF response: ((du($^{1}$HAR$_{AP}$)$_{IP}$)(($^{2}$BADEENDENE-dine$_{AP}$)$_{IP}$)$_{IU}$)
you HAVE RUBBER DUCKS$_{DEF}$-yours
'You DO have your rubber ducks.'

(10) **Control condition**
Introduction by experimenter: *Next is a picture of a girl.*
Prior NEUTRAL belief: *I don't know what the girl does.*
Visual stimuli: A girl hugging her teddy bear.
Elicitation question: *What is the girl doing?*
Potential (non PF) response: ((hun($^{2}$koser$_{AP}$)($^{2}$BAMSEN-sin$_{AP}$)$_{IP}$)$_{IU}$)
she hugs TEDDYBEAR$_{DEF}$-hers
'She is hugging her teddy bear.'

# 3. Results

## 3.1. Production of Polarity Focus

The first author coded the full sample of 460 elicitations in the five conditions for productions of PF and presence of the *jo* particles. 20% of the data were second coded, obtaining a Cohen's Kappa score of $\kappa = 0.72$, indicating substantial agreement. A third coder was used to decide in cases of disagreement.

The results show that PF is produced by all age groups (see Figure 1). While the two-year-olds produced PF in 14% of the trials ($n = 100$), the three-year-olds produced PF in 22% of the trials ($n = 100$). The four-year-olds ($n = 155$) and five-year-olds ($n = 105$) produced PF in 21% of the trials.

To investigate the development of PF production with age, we fitted a Generalized Linear Mixed Model (GLMM) of the PF productions as a count response with an upper bound, with Age as a continuous predictor, and Subjects as a random factor using a binomial error distribution and the `glmer` function of the `lme4` package (Bates et al., 2015) in R (version 4.2.2; R Core Team, 2022). The results show no effect of Age ($p = 0.192$). This suggests that the ability to produce PF overall is present already from the age of two years and that there is no significant increase in PF productions with age.

Figure 2 shows that PF was produced in all PF conditions but the control condition, suggesting that children are using this intonation pattern in appropriate contexts. In the least complex PF condition, where a positive prior belief is affirmed by the production of PF (the Pos-Aff condition), children produced PF in 28% of the trials. In the Pos-Den condition, where the production of PF involved a contradiction of a prior positive belief, PF was produced in 17% of the trials. The Neg-Den condition, where the handpuppet's prior belief was negative, and the production of PF involved a denial of a prior negative belief, was the one which elicited the highest number of PF with children producing this intonation pattern in 47% of the trials. In the most complex Inf-Bel condition, where the production of PF was relevant only if the participants had inferred that the handpuppet held a false belief which they then contradicted, PF was produced in 7% of the trials.

To investigate whether there are significant differences in PF productions between the four PF conditions, we fitted a GLM of the PF productions as a binomial response analyzed as a function of Condition as a categorical factor, using the `glm` function of the `stats` package in R (see Table 1 for a summary of the model).

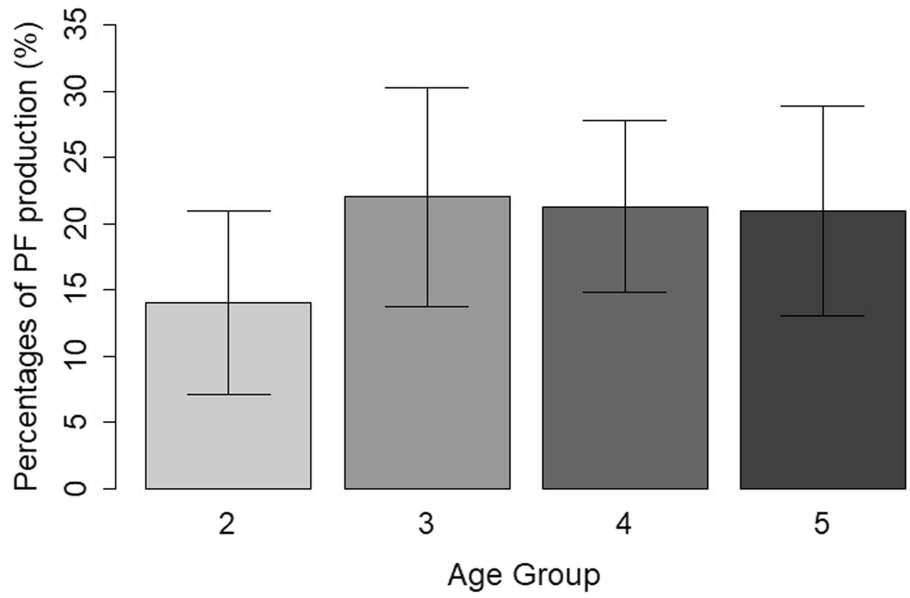



FIGURE 1
Percentages of PF productions by age group.

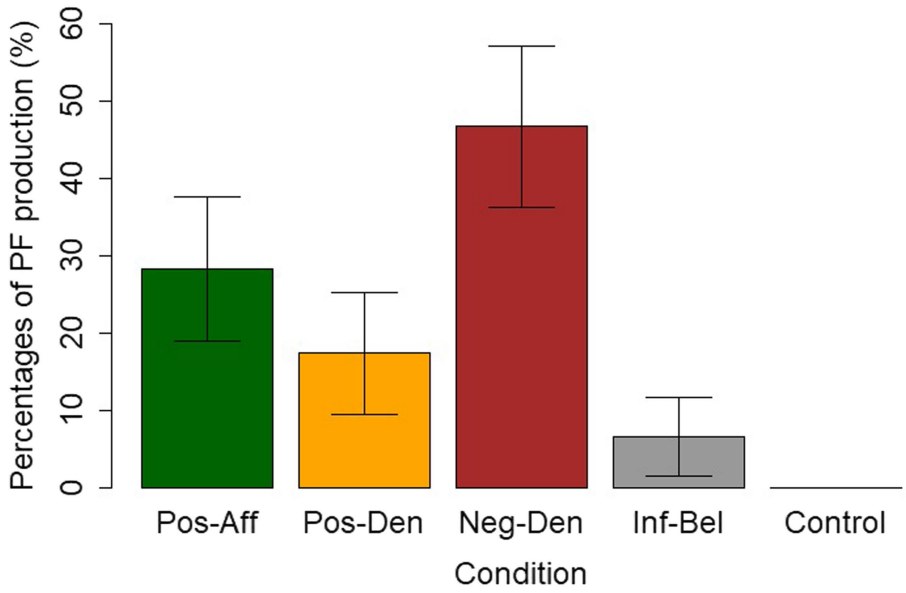



FIGURE 2
Percentages of PF productions by condition (*N*=92).

TABLE 1 Summary of GLM with PF production as a binominal response analyzed as a function of condition.

| | GLM of PF production ~ Condition | | |
|---|---|---|---|
| Condition | Odds ratios | 95% CI | *p* |
| (Intercept) | 0.07 | 0.03–0.15 | <0.001 |
| Pos-Aff | 5.65 | 2.33–15.88 | <0.001 |
| Pos-Den | 3.02 | 1.18–8.77 | 0.028 |
| Neg-Den | 12.58 | 5.34–34.85 | <0.001 |
| Observations | 368 | | |
| $R^2$ Tjur | 0.118 | | |

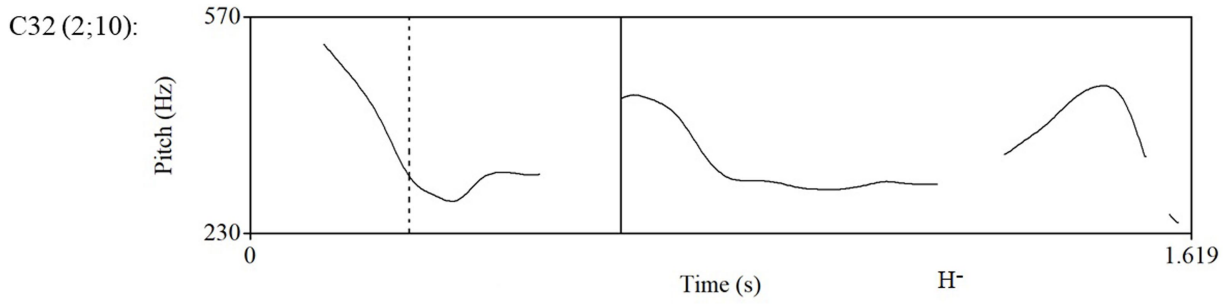
The results show that compared to the Inf-Bel condition, PF was produced significantly more often in the Pos-Aff condition ($p = <0.001$), Pos-Den condition ($p = 0.028$), and Neg-Den condition ($p = <0.001$).

Figures 3–6 below provide some examples of PF productions across conditions and age groups. Each figure consists of the fundamental frequency ($F_0$) contour of the utterance and a corresponding table with four tiers. We used the software *Praat* (version 6.2.07; Boersma and Weenink, 2022) to create the $F_0$-contours. The vertical lines in the $F_0$-contour mark intonational boundaries that correspond to the parentheses in the transcription of the $F_0$-contour given in the first tier of the table. This transcription is based on the Trondheim model (e.g., Nilsen, 1992; Fretheim, 2002) developed for analysis of Norwegian intonation. The second tier contains an annotation of the utterance's realized tones, and the two last tiers are the English translation of the utterance. The realization of PF can be seen in the $F_0$-contour as a tonal rise to an extra high tone ($H^-$) on the finite verb, followed by another tonal rise, either to a high ((H)) or an extra high ($H^-$) tone.

Figure 7 below shows the PF productions by age in each condition. First, we find PF productions in three of the four PF conditions for all age groups. The Neg-Den condition has the highest percentage of PF productions for all age groups (two-year-olds: 30%; three-year-olds: 55%; four-year-olds: 45%, and five-year-olds: 57%). While the percentages of PF productions in the Pos-Den condition are quite similar across all age groups (two-year-olds: 20%; three-year-olds: 15%; four-year-olds: 16%; five-year-olds: 19%), the PF productions in the Pos-Aff condition show an equal percentage of PF productions by three-and four-year-olds (35%) and the two-and five-year-olds (20 and 19%, respectively). Furthermore, one three-year-old produced PF in the Inf-Bel condition, but most PF productions in this condition are by the two oldest age groups, although they were not frequent overall (only amounting to six occurrences in total). Taken together, these results indicate that although the overall ability to use PF is in place at the age of two, the ability to use PF in the most complex PF condition, the Inf-Bel condition, emerges later.

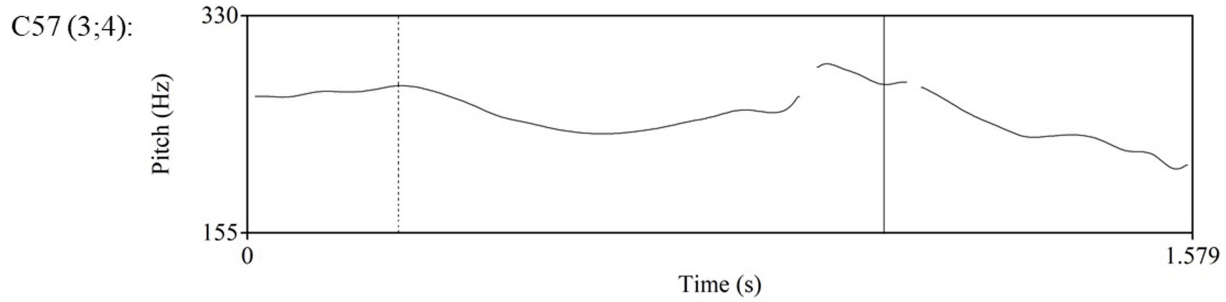
As shown by the error bars in Figure 7, there is great variance in the data, due to few observations in each condition when dividing responses into age groups. The models we present in what follows should therefore be interpreted with caution. To investigate any differences in performance of each age group in the four PF conditions, we fitted GLM models of PF productions as a binomial response analyzed as a function of the predictors Age Group and Condition, including their interaction, using the `glm` function of the `stats` package in R.[8] For each model we changed reference level for Age Group (using 5-, 4-, 3- and

8 The GLMM model that included Subjects as a random factor did not converge.



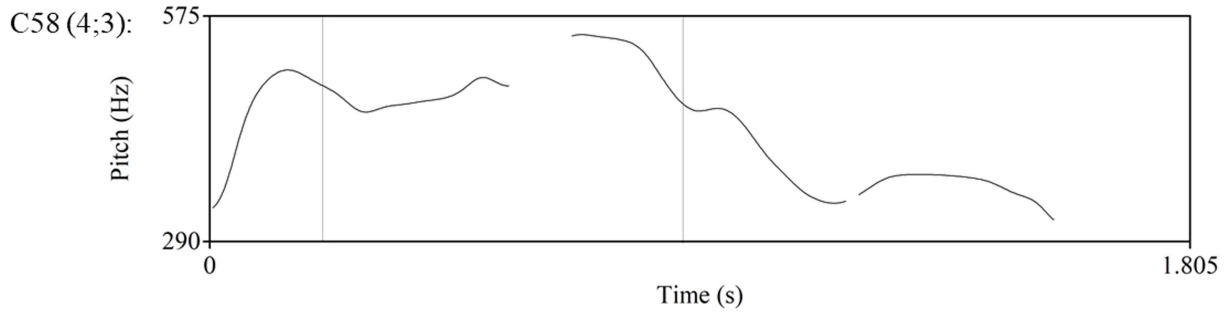

| ((han | ($^{1}$SPISER$_{AP}$)$_{IP}$) | | (($^{1}$JORDBÆR$_{AP}$)$_{IP}$)$_{IU}$) | |
|---|---|---|---|---|
| | L* | H- | L* | H- L% |
| he | EATS | | STRAWBERRIES | |
| 'He DOES eat strawberries' | | | | |

FIGURE 3
PF production by 2-year-old in the Pos-Aff condition.



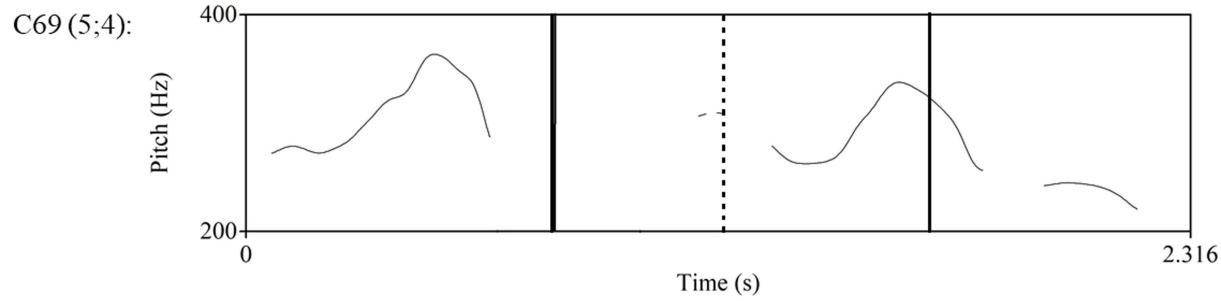

| ((henne | ($^{1}$GJØR-ikke$_{AP}$)$_{IP}$) | | ($^{1}$det$_{AP}$)$_{IU}$) | | |
|---|---|---|---|---|---|
| | L* | H- | L* | (H) | L% |
| she | DOES not | | that | | |
| 'She does NOT do that.' | | | | | |

FIGURE 4
PF production by 3-year-old in the Pos-Den condition.




| (($^{1}$den$_{AP}$) | ($^{1}$LESER$_{AP}$)$_{IP}$) | | ($^{1}$bok$_{AP}$)$_{IU}$) | | |
|---|---|---|---|---|---|
| L* (H) | L* | H- | L* | (H) | L% |
| it | READS | | book | | |
| 'It [he] DOES read a book.' | | | | | |

FIGURE 5
PF production by 4-year-old in the Neg-Den condition.




| (((1ENDER$_{AP}$)$_{IP}$)$_{IU}$) | ((du | (1HAR-jo$_{AP}$)$_{IP}$) | (1ender$_{AP}$)$_{IU}$) |
|---|---|---|---|
| L* H- L% | | L* H- | L* (H) L% |
| DUCKS | you | HAVE-PART | ducks |
| 'Ducks! You DO have ducks.' | | | |

FIGURE 6
PF production by 5-year-old in the Inf-Bel condition.

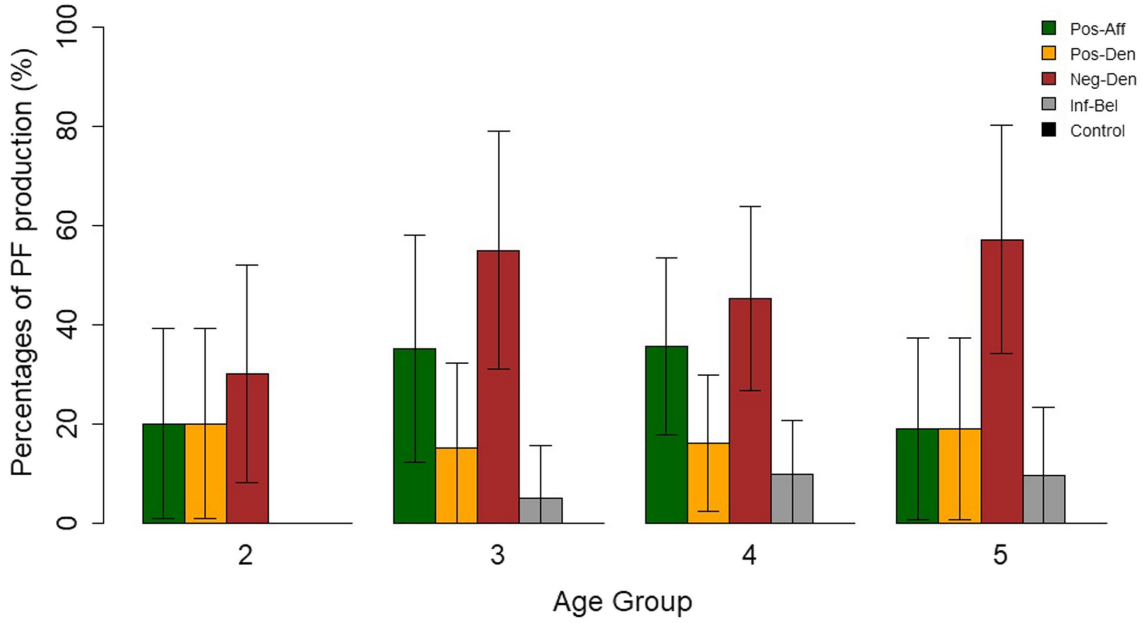



FIGURE 7
Percentages of PF productions by age group and condition.

2-year-olds), but kept the reference level for Condition constant (using the Neg-Den condition). The results show no significant differences between the age groups. There are, however, significant differences in PF productions within the age groups: Both three-, four-, and five-year-olds produced PF significantly more often in the Neg-Den condition than in the Pos-Den condition (3-year-olds: $p = 0.012$; 4-year-olds: $p = 0.017$; 5-year-olds: $p = 0.014$) and the Inf-Bel condition (3-year-olds: $p = 0.005$; 4-year-olds: $p = 0.004$; 5-year-olds: $p = 0.003$). In addition, five-year-olds produced PF significantly more often in the Neg-Den condition than in the Pos-Aff condition ($p = 0.014$). There were no significant differences in the PF productions by the two-year-olds in the Pos-Aff, Pos-Den and Neg-Den conditions.

## 3.2. The two *jo* particles

When analyzing the production data, we noticed a striking frequency of the Norwegian word form orthographically expressed as *jo*. Although we did not aim specifically at eliciting it, this word form occurs in the participant responses in 15% (68/460) of all the trials in the structured elicitation task, and in many of the cases, it co-occurs with the participants' PF productions. We therefore decided that the *jo* particles deserved more careful analyses. We first consider the productions of the sentence-initial response particle *jo* (47% of the occurrences), and then the productions of the sentence-internal pragmatic particle *jo* (53% of the occurrences). Given how little we know about the interaction between PF and the two *jo*

particles, our analyses in this section are mainly of a descriptive and qualitative character, providing the foundation for further experimental analysis.

### 3.2.1. Sentence-initial response particle *jo*

As Table 2 below shows, the response particle *jo* is produced most often in the Neg-Den condition (31% of the 92 Neg-Den trials). The following example shows a typical case where the response particle *jo* is followed by a PF utterance:

(11) Experimenter: *Next is a picture of a boy.*
Handpuppet: *I believe that the boy is not reading a book.*
C98 (2;8): $(((^{1}JO_{AP})_{IP})_{IU})$, $((han(^{1}LESER_{AP})_{IP})(^{1}bok_{AP})_{IU})$
$YES_{RESP.PART}$ he READS book
'Yes, he DOES read a book.

Table 2 shows three occurrences of the response particle *jo* in the Pos-Aff condition, a condition where *jo* should not be a relevant response since the prior belief expressed by the handpuppet is positive. One of these cases was a participant (5;1) who got excited when the handpuppet declared that he thought the boy in the upcoming picture was eating strawberries. *I don't think he is!* the participant exclaimed while waiting for the experimenter to show the picture. When the picture turned out to be a match with the handpuppet's prior positive belief (i.e., the boy in the picture is eating strawberries), the participant contradicted his own negative belief by using the response particle *jo*, and thereby made the use of *jo* relevant.

Table 3 shows the productions of the response particle *jo* in the Neg-Den condition ($N=29$). We see that in 72% of the trials participants responded using the response particle *jo* as a direct response to the puppet's prior negative belief (noFNQ), that is, without the handpuppet having to ask the elicitation question. Remember that the elicitation question was asked only if the participant did not respond to the prior belief expressed by the puppet. In the remaining 28% of the trials participants responded with *jo* in the context of a false negative question (FNQ) (i.e., *Does the boy not read a book?*). Furthermore, the response particle *jo* is followed by a PF utterance in 37% of the responses. In 28% of the responses, *jo* is followed by an utterance which is not realized with PF, and in 34% of the responses, it is used as a single word response with no succeeding utterance.

The response in (11) above was produced by a two-year-old, but, as Table 4 shows, responses consisting of the response particle *jo* followed by a PF utterance are produced by participants in all age groups. The combination of *jo* and an utterance without PF, as well as *jo* as a single word response also occur in all age groups. In other words, sentence-initial *jo* seems to be available already from two years of age, also felicitously co-occurring with PF from this early age.

TABLE 2 Productions of sentence-initial response particle *jo* by condition.

| Conditions | *Jo* |
|---|---|
| Pos-Aff | 3 |
| Pos-Den | - |
| Neg-Den | 29 |
| Inf-Bel | - |
| Control | - |
| Total | 32 |

TABLE 3 Productions of sentence-initial response particle *jo* in the Neg-Den condition ($N=29$).

| Neg-Den cond. | FNQ | noFNQ |
|---|---|---|
| Resp.part *Jo* | 28% | 72% |
| *Jo* + PF | 3% | 34% |
| *Jo* + noPF | - | 28% |
| Simple *Jo* | 24% | 10% |

### 3.2.2. Sentence-internal pragmatic particle *jo*

In the utterance in Figure 6 above, repeated below as (12) for convenience, we saw an example of a PF utterance used as a response in the Inf-Bel condition. This utterance also included a sentence-internal *jo*:

(12) Handpuppet: *I wish I had something to play with while taking a bath!*
C69 (5;04): $(((^{1}ENDER_{AP})_{IP})_{IU})$, $((du(^{1}HAR\text{-}jo_{AP})_{IP})(^{1}ender_{AP})_{IU})$
DUCKS, you HAVE-$_{PART}$ ducks
'Ducks, you DO have ducks (remember).'

Table 5 shows the productions of sentence-internal *jo* in the five conditions. The highest number appears in the Inf-Bel condition (44%), followed by the Neg-Den condition (25%), the Pos-Aff condition (17%), the Pos-Den condition (8%), and the Control condition (6%). It further shows that sentence-internal *jo* is produced both in combination with PF and in utterances not carrying PF in all four PF conditions. The occurrences in the Control condition might seem odd, given that the use of PF is not relevant here. However, although sentence-initial *jo* naturally co-occurs with PF, addressing the polarity of a proposition expressed is not part of the semantics of this pragmatic particle. It can therefore felicitously be used in utterances produced in a neutral (or ignorant) context, such as our Control condition.

Table 6 shows the productions of sentence-internal *jo* across age groups. While there are no productions among the two-year-olds, we find occurrences in the three other age groups (3-year-olds: 28%; 4-year-olds: 31%; and 5-year-olds: 42%). In addition, three-, four-, and five-year-olds produced sentence-internal *jo* both in combination with PF and in utterances without any realized PF.

In the responses from the elicitation task we also find some non-PF utterances (i.e., utterances that do not meet the intonational criteria for PF of having both a focally accentuated polarity carrier and an additional succeeding

TABLE 4 Productions of the sentence-initial response particle *jo* by age group ($N=32$).

| | 2-year-olds | 3-year-olds | 4-year-olds | 5-year-olds |
|---|---|---|---|---|
| Resp.part *Jo* | 31% | 31% | 22% | 16% |
| *Jo* + PF | 9% | 22% | 9% | 6% |
| *Jo* + noPF | 9% | 3% | 9% | 3% |
| Simple *Jo* | 13% | 6% | 3% | 6% |

TABLE 5 Productions of the Norwegian sentence-internal pragmatic particle *jo* by condition (*N*=36).

| | Pos-Aff | Pos-Den | Neg-Den | Inf-Bel | Control |
|---|---|---|---|---|---|
| Pragm. part *Jo* | 17% | 8% | 25% | 44% | 6% |
| PF w/*jo* | 8% | 6% | 19% | 17% | - |
| noPF w/*jo* | 8% | 3% | 6% | 28% | 6% |

TABLE 6 Productions of the Norwegian sentence-internal pragmatic particle *jo* by age group (*N*=36).

| | 2-year-olds | 3-year-olds | 4-year-olds | 5-year-olds |
|---|---|---|---|---|
| Pragm. part *Jo* | - | 28% | 31% | 42% |
| PF w/*jo* | - | 19% | 17% | 14% |
| noPF w/*jo* | - | 8% | 14% | 28% |

accentuation) that seem to have a pragmatic effect similar to a PF utteranrce. Consider the example in Figure 8 from one of the participants produced in the Inf-Bel condition after the handpuppet had said *I wish I had something to play with while taking a bath!*

The utterance in Figure 8 has a focally accentuated finite verb, but since there is no following accentuation later in the utterance, it is not considered PF utterance. However, a sentence-internal *jo* follows as an unaccented syllable in the tail of the rising tone. The utterance seems to have a similar effect as a PF response would have had: the participant signals that there is some sort of conflict between the handpuppet's belief (that he does not have any toys to play with in the bath) and the current state of affairs (that he owns rubber ducks).

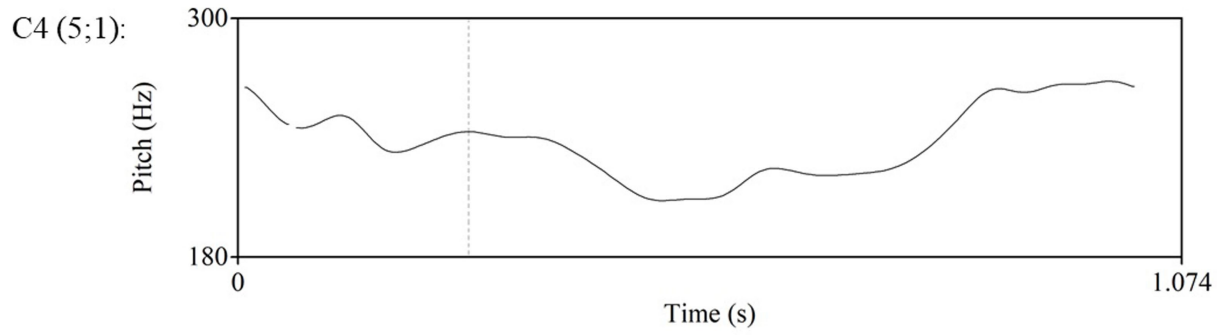
## 4. Discussion

The aim of our study was to investigate whether Norwegian-speaking children aged two to five years can produce intonation utterances realized with 'Polarity Focus' (PF), and if so, whether we would find a gradual development of their productions in contexts of increasing complexity. While we found productions of PF in all age groups tested and only in felicitous contexts, our hypothesis of a gradual development was only partially supported by our data.

Overall, our findings show that children produce PF from as early as age two. We take this to suggest that they are also (at some level) able to evaluate the truth or falsity of a proposition and attribute a contextually available proposition to their interlocutor from around this age. Furthermore, our findings show that young children can express their affirmation or denial of this truth-conditional content by intonational means. Already from the age of two and onwards, children seem to use intonation naturally and efficiently as a communicative device and in this case specifically to signal epistemic vigilance toward an attributed propositional content. This is also likely to involve an intention to modify their interlocutor's epistemic state.

As expected, the ability to use PF to express the denial of an inferred false belief seems to arise around four years of age. However, the percentage of PF productions in this condition was overall low (7%). Since this is the most complex condition of our design and only expected to be mastered by the older children, the low percentage of PF productions was not surprising. Our findings align with previous findings in the Theory of Mind literature where the ability to linguistically express an understanding of others' false beliefs manifests around four years of age (*cf.* Wellman et al., 2001).

Our finding that PF was produced by two-year-olds in both the Pos-Aff, the Pos-Den and the Neg-Den conditions support previous research that show that children are able to both reject false and accept true statements before their second birthday (Lyon et al., 2013), and that two-year-olds spontaneously correct assertions they believe to be false (Pea, 1982). In fact, it was the Neg-Den condition




| | |
|---|---|
| ((det | (([1]HAR-du-jo$_{AP}$)$_{IP}$)$_{IU}$) |
| | L* H- H% |
| that | HAVE-you-PART |
| 'You have that, you know.' | |

FIGURE 8
F0-contour of 'non-PF utterance' with similar pragmatic effect as a PF utterance.

that had by far the highest percentage of PF productions in all age groups. This finding is surprising since the literature suggests that an increasing number of negations increases the complexity of the test conditions (*cf.* Just and Carpenter, 1971). Although this may be the case at a general level, we see no evidence of this in our data: Even two-year-olds master the more complex context where they need to contradict a prior negative belief. This result is in line with Pea (1982), who shows that the ability to correct false statements appears prior to their expressing agreement with true statements. Intonation, and in this case PF, seems to offer young children an easily accessible linguistic strategy for communicating their attitude toward a propositional content, enabling even children as young as two years to express this higher-level metarepresentational content without having to verbalize it.

This high percentage of PF productions in the Neg-Den condition suggests that it was the most natural context for PF production in our task. A growing body of research shows that from very early on children monitor the reliability of the information communicated (Gliga and Csibra, 2009; Koenig and Woodward, 2010; Sperber et al., 2010). It is possible that signaling a denial or an opposing opinion might be more socially important than signaling an endorsement. It could also be that PF is more frequently used by adults in contexts like the Neg-Den condition, and therefore possibly more familiar to children. We know from the study by Turco et al. (2014) that adults produced Verum Focus in 70% of the 'polarity correction' contexts where the verbal stimuli used was similar to our Neg-Den condition, involving a mismatch in form of a false negative statement about what was depicted in the visual stimuli. Future research should investigate experimentally how Norwegian-speaking adults use PF, focusing on the different contexts for eliciting PF and in what ways they differ.

We further observed that the Neg-Den condition also provided a natural context for responding with the sentence-initial response particle *jo*, often in combination with PF (such as in (11) above). While according to Noveck et al. (2021) the accurate but surprisingly fast *si* response by the four-year-olds in their study suggested that these children did not aim at modifying the epistemic state of their interlocutor when responding with *si*, our results suggest otherwise. First, they indicate that the ability to produce pragmatically felicitous responses using the response particle *jo* could be present as early as two years of age. Given what we know from the developmental literature of children's ability to evaluate the truth value of propositional content at such early age (Pea, 1982; Lyon et al., 2013), together with some level of perspective taking (O'Madagain and Tomasello, 2021), it seems likely that, if two-and three-year-olds produce *jo* accurately, they have by the age of four developed a pragmatic maturity that goes beyond relying merely on the minimal semantic representation of the particle. Second, children younger than four years showed mastery of PF production and especially in the Neg-Den condition, a context highly similar to the Negative-Si condition used by Noveck et al. (2021) to elicit the response particle. Furthermore, in our study, children younger than four years spontaneously and felicitously produced the combination of the sentence-initial response particle *jo* and PF. Together, this mastery of both PF and the response particle *jo* at such early age, we believe speak against a limited pragmatic competence involved in the use of this particle at four years of age. To gain a deeper understanding of how the response particle *jo* and PF are related and what the use of them separately and in combination can reveal about children's developing pragmatic abilities, future research should address this relationship directly, using different approaches and methodologies and a broader set of context types to elicit the two phenomena.

Our data also included participant responses that, although realized without PF, seemed to have a similar pragmatic effect. In the example in Figure 8 above, we discussed how this effect could be due to the presence of a sentence-internal *jo*, which often involves some sort of oppositional feature. Berthelin and Borthen (2019) argue that the procedural meaning encoded by *jo* involves an instruction to the hearer to interpret the proposition expressed as mutually manifest to speaker and hearer, and to use the proposition expressed as a premise for deriving and supporting a contextual implication. As they (2019, p. 25) point out: "*jo* is a useful tool when speakers suspect that the hearer will not accept the information they are communicating." This is also the case for PF. Just like sentence-internal *jo*, PF can be used when a speaker needs to convince her interlocutor of the epistemic status of the proposition expressed and of the conclusions that can be drawn from it. Future research should also investigate the relationship between the sentence-internal particle *jo* and PF in more detail. If the two phenomena are closely related, in what ways do they differ, and what triggers the use of them in combination?

We have suggested that utterances carrying PF have an affinity with echoic use in the relevance-theoretic sense (Wilson, 2012): PF utterances involve both an attribution of the proposition expressed, and they enable the speaker to convey her attitude to this proposition in the form of a denial or an affirmation. Although this is a rather simple form of echoic use, the attitude being explicitly conveyed, our study has shown that children master such echoic uses from a very early age. This has potential implications for theoretical accounts of the development of other, more complex forms of echoic use such as verbal irony, which is characterized by the speaker tacitly echoing and expressing a dismissive attitude to an attributed thought (Wilson, 2012). These uses have been shown to have a protracted development (Falkum and Köder, 2020) with some comprehension abilities emerging around the age of six years. Our results show that the ability to express an endorsing or dismissive attitude to an attributed thought (expressed explicitly in the context in the simplest cases) emerges much earlier. In this way, intonational competence, and more specifically the ability to use PF appropriately, could be seen as a precursor to ironical uses.

Finally, we would like to mention some caveats. First, we have claimed that PF production involves attribution of the thought that is affirmed or denied by the use of PF. However, the design of the verbal stimuli in our first three PF conditions (Pos-Aff, Pos-Den, Neg-Den) makes it difficult to tease apart this ability from the ability to metarepresent a thought (without having to attribute it), since the handpuppet explicitly expresses his prior beliefs. There is solid evidence that four-and five-year-olds can attribute thoughts. One possibility then is that the two-and three-year-olds do not attribute the thought they are affirming or denying, but are merely metarepresenting a contextually available thought. However, this analysis leaves open the question why a speaker would produce PF in the first place: The informative intention of a speaker who uses PF is to convey her affirmation or denial of a metarepresented thought. Why would she convey this if not to modify her interlocutor's epistemic state (which does involve thought attribution)? The production of PF does not make sense in a context where the speaker

merely metarepresents the thought without attributing it to someone. If two-and three-year-olds did not have this ability (and as a consequence they are not aiming at modifying their interlocutor's epistemic state), we should expect them to produce less PF overall than the older children, simply because they would not experience relevant situations for the production of PF. However, we find no significant differences in PF productions between the age groups in our study. It seems likely, therefore, that the ability to attribute thoughts is also involved in the felicitous production of PF, and that this ability is present already from the age of two years.

Second, our experimental setting posed some challenges especially for the youngest participants. Although the experimenter made an effort to make the conversation as natural as possible, some of the two-year-olds had trouble adapting to the experimental setting or did not feel familiar enough with the experimenter (even though all participants who wanted to bring a familiar caretaker were given the opportunity to do so), and refused to speak. This could have masked their intonational competence. Finally, the use of production data as a source of evidence for pragmatic competence requires an interpretation of children's communicative intention, which is inevitably speculative (Zufferey, 2020). Moreover, production data are often thought to underestimate children's performance compared to comprehension data. However, since the conditions in our structured elicitation task are specifically designed to elicit PF, less is left to speculation compared to spontaneous productions in unstructured contexts. We also believe that for the study of early pragmatic development, production data provide a valuable source of insight, especially because controlled comprehension experiments may not be feasible with children in the youngest age groups. However, our conclusions in this paper inevitably rest on our interpretation of the production data.

## 5. Conclusion

Our study provides the first experimental evidence that Norwegian-speaking children are able to produce intonation utterances realized with 'Polarity Focus' from an early age. We suggest that the mastery of the production of PF, as well as their ability to produce the *jo* particles in appropriate contexts, can be seen as an early linguistic manifestation of the cognitive abilities for the attribution of thoughts and epistemic vigilance toward propositional content.

At a more general level, our study provides insight into the role of intonation as part of a broader pragmatic competence. An overarching aim was to start filling a gap in the literature by combining suprasegmental phonology and cognitive pragmatic theory. We provide experimental evidence for the pragmatic functions of intonation, which in the case of PF allows speakers to communicate a positive or negative attitude toward a metarepresented proposition. Our exploratory analyses of the *jo* particles also contribute some insight into children's developing metarepresentational abilities.

We believe to have shown that studying intonational production can be useful as a window into children's pragmatic competence. Although our results did not fully support the developmental hypothesis, they provide evidence of the intonational productions of children aged two to five years and a piece of information about their developing pragmatic competence which is currently missing in the literature. We hope to see many more studies of children's intonational competence during this crucial developmental period in the coming years.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by NSD - Norwegian Center for Research Data. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

LH: conceptualization, experimental design, data collection, coding of the data, writing–original draft preparation, writing–review and editing, visualization, data analysis, and project administration. IF: conceptualization, experimental design, writing–review and editing, data analysis, and supervision. All authors contributed to the article and approved the submitted version.

## Acknowledgments

We are grateful to the children who participated in our experiment, and to the parents and kindergartens for making data collection possible. We would like to thank Thorstein Fretheim for second coding the material and discussing all imaginable (and unimaginable) aspects of Norwegian intonation and, in particular, the phenomenon of Polarity Focus, Randi A. Nilsen for being the third coder and for discussing possible experimental designs, Franziska Köder for invaluable help with the statistical analyzes, and Camilo Rodriguez Ronderos and Mary Beth Neff for commenting on previous drafts of the paper. We thank the two reviewers for insightful comments and suggestions.

## 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.

## Publisher's note

# References


Bates, D., Maechler, 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

Berger, F., and Höhle, B. (2012). Restrictions on addition: children's interpretation of the focus particles Auch 'also' and nur 'only' in German. *J. Child Lang.* 39, 383–410. doi: 10.1017/S0305000911000122

Berthelin, S. R., and Borthen, K. (2019). The semantics and pragmatics of Norwegian sentence-internal jo. *Nordic J. Linguist.* 42, 3–30. doi: 10.1017/S0332586519000052

Blakemore, D. (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.

Blakemore, D. (2002). *Relevance and Linguistic Meaning: The Pragmatics and Semantics of Discourse Markers*: Cambridge: Cambridge University Press.

Boersma, P., and Weenink, D. (2022). *Praat: Doing Phonetics by Computer (Version 6.2.07)*. Available at: praat.org (Accessed January 28, 2022).

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

Dimroth, C., Schimke, S., and Turco, G. (2018). Focusing functional elements: affirmative particles and Verum focus in first language Acquisition of German. *Lang. Acquis.* 25, 268–283. doi: 10.1080/10489223.2017.1294594

Falkum, I. L., and Köder, F. (2020). The acquisition of figurative meanings. *J. Pragmat.* 164, 18–24. doi: 10.1016/j.pragma.2020.04.007

Fretheim, T. (2002). *Intonation as a Constraint on Inferential Processing. Paper Presented at the Speech Prosody 2002*. France: Aix-en-Provence. Available at: http://sprosig.org/sp2002/pdf/fretheim.pdf

Fretheim, T. (2017). "The pragmatics of 'Yes' and 'No'" in *Mouton Series in Pragmatics. Pragmatics at its Interfaces*. ed. S. Assimakopoulos, vol. *17* (Berlin, New York: de Gruyter Mouton), 175–200.

Fretheim, T., and Nilsen, R. A. (1993). "The Norwegian broad-focus avoidance strategy," in *Nordic Prosody VI: Papers from a Symposium, Stockholm, August 12–14, 1992*. eds. B. Granström and L. Nord (Stockholm: Almqvist & Wiksell International), 49–64.

Gliga, T., and Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychol. Sci.* 20, 347–353. doi: 10.1111/j.1467-9280.2009.02295.x

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:n/a. doi: 10.1111/lnc3.12310

Gussenhoven, C. (1983). Focus, mode and the nucleus. *J Linguist.* 19, 377–417. doi: 10.1017/S0022226700007799

Just, M. A., and Carpenter, P. A. (1971). Comprehension of negation with quantification. *J. Verbal Learn. Verbal Behav.* 10, 244–253. doi: 10.1016/S0022-5371(71)80051-8

Kehoe, M. (2013). *Development of Prosody and Prosodic Structure*. Hauppauge: Nova Science Publishers, Incorporated.

Koenig, M. A., and Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: possible mechanisms. *Dev. Psychol.* 46, 815–826. doi: 10.1037/a0019664

Kristoffersen, K. E., and Simonsen, H. G. (2012). *Tidlig Språkutvikling Hos Norske Barn: MacArthur-Bates Foreldrerapport for Kommunikativ Utvikling*. Oslo: Novus.

Leslie, A. M. (1987). Pretense and representation: the origins of "theory of mind.". *Psychol. Rev.* 94, 412–426. doi: 10.1037/0033-295X.94.4.412

Lyon, T. D., Quas, J. A., and Carrick, N. (2013). Right and righteous: Children's incipient understanding and evaluation of true and false statements. *J. Cogn. Dev.* 14, 437–454. doi: 10.1080/15248372.2012.673187

Matthews, D., Lieven, E., Theakston, A., and Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young childrens use of referring expressions. *Appl. Psycholinguist.* 27, 403–422. doi: 10.1017/S0142716406060334

Nilsen, R. A. (1992). *Intonasjon i Interaksjon: Sentrale Spørsmål i Norsk Intonologi, (PhD Dissertation)*. University of Trondheim, Trondheim, Norway.

Noveck, I. A., Petit, N., Tian, Y., and Turco, G. (2021). Revealing pragmatic processes through a one-word answer: when the French reply Si. *J. Mem. Lang.* 120:104245. doi: 10.1016/j.jml.2021.104245

O'Madagain, C., and Tomasello, M. (2021). Joint attention to mental content and the social origin of reasoning. *Synthese* 198, 4057–4078. doi: 10.1007/s11229-019-02327-1

Pea, R. D. (1982). Origins of verbal logic: spontaneous denials by two-and three-year olds. *J. Child Lang.* 9, 597–626. doi: 10.1017/S0305000900004931

Peppé, S., and Wells, B. (2014). "Speech Prosody," in *Encyclopedia of Language Development*. eds. P. J. Brooks and V. Kempe (Thousand Oaks, CA: Sage Publications), 585–590.

Pouscoulous, N., Noveck, I. A., Politzer, G., and Bastide, A. (2007). A developmental investigation of processing costs in Implicature production. *Lang. Acquis.* 14, 347–375. doi: 10.1080/10489220701600457

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Romøren, A. S. H., and Chen, A. (2021). The acquisition of prosodic marking of narrow focus in central Swedish. *J. Child Lang.* 49, 213–238. doi: 10.1017/S0305000920000847

Scott, K. (2021). "Contrastive stress in English: meaning, expectations and ostension," in *Pragmatics and Beyond New Ser. Beyond Meaning*. eds. T. Wharton, L. D. Saussure and E. Ifantidou, vol. *324* (Amsterdam, Netherlands: John Benjamins)

Sperber, D., ClÉMent, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., et al. (2010). Epistemic Vigilance. *Mind Lang.* 25, 359–393. doi: 10.1111/j.1468-0017.2010.01394.x

Sperber, D., and Wilson, D.. (1986/1995) *Relevance: Communication & Cognition 2nd*. Malden, Massachusetts: Blackwell Publishers.

Tomlinson, J. M., Gotzner, N., and Bott, L. (2017). Intonation and pragmatic enrichment: how intonation constrains ad hoc scalar inferences. *Lang. Speech* 60, 200–223. doi: 10.1177/0023830917716101

Turco, G., Braun, B., and Dimroth, C. (2014). When contrasting polarity, the Dutch use particles, Germans intonation. *J. Pragmat.* 62, 94–106. doi: 10.1016/j.pragma.2013.09.020

Turco, G., Dimroth, C., and Braun, B. (2013). Intonational means to mark Verum focus in German and French. *Lang. Speech* 56, 461–491. doi: 10.1177/0023830912460506

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Wells, B., Peppé, S., and Goulandris, N. (2004). Intonation development from five to thirteen. *J. Child Lang.* 31, 749–778. doi: 10.1017/S030500090400652X

Wharton, T. (2012). "Pragmatics and prosody," in *The Cambridge Handbook of Pragmatics*. eds. K. Allan and K. M. Jaszczolt (Cambridge: Cambridge University Press), 567–584.

Wharton, T. (2020). "Acquiring prosody," in *Developmental and Clinical Pragmatics*. eds. K. P. Schneider and E. Ifantidou, vol. *13* (Berlin: Mouton de Gruyter).

Wieman, L. A. (1976). Stress patterns of early child language. *J. Child Lang.* 3, 283–286. doi: 10.1017/S0305000900001501

Wilson, D. (2011). "The conceptual-procedural distinction: past, present and future," in *Procedural Meaning: Problems and Perspectives*. eds. V. Escandell-Vidal, M. Leonetti and A. Ahern (Bingley: Emerald), 3–31.

Wilson, D. (2012). "Metarepresentation in linguistic communciation," in *Meaning and Relevance*. eds. D. Wilson and D. Sperber (Cambridge: Cambridge University Press), 230–258.

Wilson, D. (2016). Reassessing the conceptual–procedural distinction. *Lingua* 175-176, 5–19. doi: 10.1016/j.lingua.2015.12.005

Wilson, D., and Wharton, T. (2006). Relevance and prosody. *J. Pragmat.* 38, 1559–1579. doi: 10.1016/j.pragma.2005.04.012

Zufferey, S. (2020). "Pragmatic development in a first language: an overview," in *Developmental and Clinical Pragmatics*. eds. K. P. Schneider and E. Ifantidou (Berlin/Boston: De Gruyter Mouton), 33–60.

frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 20 April 2023
DOI 10.3389/fcomm.2023.1125575



OPEN ACCESS

EDITED BY
Ryoko Sasamoto,
Dublin City University, Ireland

REVIEWED BY
Lucy Simmonds,
Flinders University, Australia
Alison Hall,
De Montfort University, United Kingdom

*CORRESPONDENCE
Kate Scott
✉ kate.scott@kingston.ac.uk

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

RECEIVED 16 December 2022
ACCEPTED 28 March 2023
PUBLISHED 20 April 2023

CITATION
Scott K (2023) Nutritional labeling, communication design, and relevance. *Front. Commun.* 8:1125575. doi: 10.3389/fcomm.2023.1125575



# Nutritional labeling, communication design, and relevance

Kate Scott*

Kingston School of Art, Kingston University, Kingston upon Thames, United Kingdom

In this paper, I use relevance theory to explain the relative effectiveness of three different nutrition labeling systems in communicating information and influencing consumer food choices. Facts Up Front [also known as Reference intake (RI) or Guideline Daily Amounts (GDA)], traffic light systems, and warning labels present nutritional information in different front of pack (FOP) formats. Research into the effectiveness of these systems shows that warning labels improve consumers' ability to identify unhealthy products, compared with both Facts Up Front and traffic light systems. Warnings and traffic light systems perform equally well, however, when participants are asked to identify the most healthful product. I demonstrate how these findings can be explained in terms of the processing effort and inferential steps required from the consumer when accessing relevant contextual assumptions and deriving relevant implications in decision-making contexts. That is, I show how the success of the various labeling systems is linked to their relevance in the context of interpretation. This analysis illustrates the explanatory power of relevance theory in relation to visual communication and has implications for communication design and policy more generally.

KEYWORDS

**pragmatics, relevance theory, communication design, labeling, relevance-theoretic analyses**

## 1. Introduction

The effectiveness of the communications strategies of governments and advisory bodies can influence the health-related behavior of the public (Hornik, 2002; Wakefield et al., 2010). One area in which many governments legislate and/or provide guidance and recommendations is food and drink labeling. Policies around food packaging and the presentation of nutritional information vary by region and country. There are several formats for displaying nutritional information on food packaging, and there is a wealth of research into how these systems perform, both in terms of conveying information and changing consumer behavior. However, explanations as to why some systems yield better outcomes than others remain general. For example, Temple (2020; p. 5) discusses the apparent effectiveness of two of the systems and concludes that the "most likely reason for this is that these designs are fairly easy for shoppers to understand." In this article, I use a pragmatic framework to analyze the interpretative processes that consumers go through when interpreting a label. This then allows us to unpack what "easy for shoppers to understand" might mean in terms of the cognitive processes involved in reaching an interpretation of the nutritional information. As the various labeling systems present the same basic information in different ways, they can be used as a test case for the application of pragmatic principles in communication design.

Pragmatics is the study of communication in context. Relevance-theoretic pragmatics (Sperber and Wilson, 1995; Carston, 2002; Wilson and Sperber, 2012) offers a framework for understanding how intentional acts of communication are interpreted. As such, it is well-placed to provide insight into why some nutritional labeling systems are more effective in terms of conveying information and influencing behavior than others. By applying the principles of relevance to the interpretation of labels as communicative devices, we can compare the interpretative routes that users take when they process this information. Effective communication is not just about what information is included in a message, but also about how that information is presented. An understanding of the interpretative processes which underlie consumers' engagement with nutritional labeling should feed into both design practice and communications policy decisions in the future, and it paves the way for ideas from pragmatics to inform future work within communication design.

I start in the next section by outlining the key aspects of the relevance-theoretic approach to cognition and communication. The assumptions and principles presented here underpin the analyses and discussions that follow. In Section 3, I outline the main food labeling systems that are currently in use, and I then give an overview of the main findings of research into the effectiveness of these systems in Section 4. In Section 5, I bring these ideas together and present a relevance-theoretic analysis of the labeling systems, demonstrating how differences in interpretation can be tied to differences in the design of the systems. In Section 6, I discuss some implications of this analysis, with a focus on how communication design and policy might be informed by our understanding of pragmatics and utterance interpretation.

## 2. Relevance and communication design

Relevance theory (Sperber and Wilson, 1995; Carston, 2002; Wilson and Sperber, 2006, 2012) is a framework for understanding how communicative acts (including utterances) are interpreted. At its heart are two core principles, one relating to human cognition and the other to communication. According to the cognitive principle of relevance, human cognition is geared to the maximization of relevance. An input will be relevant to an individual if it leads to cognitive effects. Cognitive effects are changes in our cognitive environment, and we can think of these as changes to the assumptions that we hold. An input might be relevant because it causes us to strengthen an assumption that we already hold. It may be relevant because it contradicts an assumption that we hold and leads us to eliminate it. Finally, an input may be relevant because it combines with an assumption that we hold to yield a new assumption that was previously unavailable to us.

Relevance is a matter of degree, and some inputs will be more relevant than others. The more cognitive effects that an input leads to (all other things being equal), the more relevant that input will be. However, processing inputs and deriving cognitive effects takes mental effort, and the more effort involved, the less relevant that input will be (again, all other things being equal). The relevance of an input is also relative to the context in which it is processed, and it is specific to the individual who is processing it. Something that is highly relevant for one person, may have little relevance for another.

According to the communicative principle of relevance, ostensive acts of communication carry with them, as part of their meaning, a presumption of their own optimal relevance. That is, when information is communicated intentionally and overtly, the addressee can assume that the communicator intended the message to be optimally relevant. The definition of optimal relevance is given in (1).

(1) (a) The ostensive stimulus is relevant enough to be worth the audience's processing effort, and (b) it is the most relevant one compatible with the communicator's abilities and preferences (Wilson and Sperber, 2006; p. 612).

This characterization of optimal relevance and the communicative principle of relevance combine to give us the relevance-theoretic comprehension procedure, given in (2).

(2) Follow a path of least effort in computing cognitive effects: Test interpretive hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility. Stop when your expectations of relevance are satisfied (or abandoned) (Wilson and Sperber, 2006; p. 613).

This framework for understanding how utterances (and other ostensive acts of communication) are interpreted has significant consequences for communication design. To interpret a message, the audience must access contextual assumptions that can combine with the input in a way that yields cognitive effects. Communicators therefore need to make predictions about the assumptions that their intended audience will hold and how strongly they will hold them. It will, for example, be much harder to change behavior if that behavior is based on assumptions that are held with a high degree of confidence. Furthermore, communication is likely to be unsuccessful if the information included in a message cannot combine with an assumption that the intended audience already holds. The task of predicting the assumptions of an audience is further complicated if there are no definite addressees or if the message is intended for a mass audience. A communicator may not know exactly who the message will reach and what assumptions they might hold. Public service announcements may be intended to communicate with a large and diverse group of people, all of whom may bring different assumptions to their interpretation.

A further consequence of this model of utterance interpretation is that the relevance of a message depends not just on the information that it includes but also on the ease with which the audience can access and process this information. Processing effort, and hence relevance, is affected by the accessibility of the information itself (Can it be clearly read? Is it written in a language that the audience understands? Does it use vocabulary that the audience members are familiar with? How linguistically or logically complicated is the information? etc.). The processing effort demanded of the audience will also be affected by the accessibility of the contextual assumptions

with which the information interacts to yield cognitive effects. Assumptions that are accessed frequently or which have been accessed recently will be more accessible than those that are rarely part of an individual's interpretation processes. The more effort that is demanded from the audience, the less relevant the message will be, and, as allowed for in the relevance-theoretic comprehension procedure in (2), if put to too much effort, the audience member may abandon the search for relevance altogether.

The principles of relevance and the definitions that underpin them provide a framework for understanding how we process and interpret new information. Relevance is comparative, and we can understand differences in the relevance of inputs in terms of the processing effort that they demand and the cognitive effects to which they lead.

Various studies have considered the role of visual and multimodal communication from the perspective of relevance theory. Forceville (2014; p. 67) has argued that relevance theory 'allows for the systematic analysis of all forms of communication in all (combinations of) modes in all media' and demonstrates the potential of this approach in his analyses of logos, advertising, political cartoons, and comics (Forceville, 2020). In an analysis of the front covers of political magazines, Tseronis (2018) uses relevance theory to demonstrate that multimodal cues in images not only attract the attention of an audience but also play a role in the communication of an argument. Relevance theory has also been used to demonstrate how the visual design of text plays a role in the communication of meaning. Sasamoto et al. (2017; p. 427) show that the "multi-colored, and highly visible, intra-lingual captions" added to some Japanese television programmes are "deliberately used to influence viewers' interpretations." Both Sasamoto and O'Hagan (2020) and Scott and Jackson (2020) consider the role that the visual appearance of text plays in the interpretation of written utterances and conclude that stylistic decisions can be used to guide the audience to an intended interpretation. I build on this work here, using insights from relevance theory to explain the patterns that we find in the effectiveness of different food product labeling systems.

# 3. Nutritional labeling and consumer perception

## 3.1. An overview of labeling policies and systems

Restrictions and requirements for nutritional labeling on food and drink products vary according to the country in which the product will be sold. Some form of nutritional information is often required by law on all pre-packaged foods, and this most often appears on the back of packaging. Regulation around front of pack (FOP) labeling varies more widely and is often voluntary. For example, in the European Union, producers must provide a nutritional declaration in a specific format, but they may also repeat that information for certain nutrients (energy, fat, saturates, sugar and salt) on the front of the food packaging (European Union, 2/11/22). In Chile, warnings must be included as part of the FOP packaging when the product exceeds a recommended limit for certain key nutrients.

According to Hersey et al. (2013) front of pack labeling falls into two main categories: nutrient specific systems and summary systems. Nutrient specific systems provide information about various key nutrients in the product. Summary systems, on the other hand, "use an algorithm to provide an overall nutritional score" (Hersey et al., 2013; p. 2). This summary may take the form of an endorsement logo indicating that the product satisfies certain requirements, or it may be a rating system of some sort, such as the Guiding Star system which rates products as "good," "better," or "best" (Guiding Stars Licensing Company) by awarding them one, two, or three stars. Nutri-Score is a summary system used in several EU countries. Products are given a rating of A to E, based on nutritional value. In a systematic review of studies into food labeling systems, Hersey et al. (2013; p. 13) conclude that "consumers more easily identify healthier foods using nutrient-specific schemes compared with the summary systems."

Hodgkins et al. (2012) propose that labeling systems can be divided into three sub-categories based on how much direction they give the consumer. They may be directive, semi-directive, and non-directive. In the analysis that follows, I look at research that compares the effectiveness of labeling systems from across this three-way categorization and explain the results using relevance-theoretic assumptions about how we interpret ostensive stimuli. A brief introduction to the three categories and the schemes which fall into them is therefore useful at this stage.

## 3.2. Directive systems

Directive labels make direct claims about the healthfulness (or otherwise) of a product and the claims are usually endorsed by a third party such as a government, charity, or regulating body. Some directive labels provide summaries, indicating that a food has been classified as meeting a certain overall standard. Others may provide direct information about one or more nutrient. Nutrition-specific directive labels make general claims ("low in fat," "high in sugar") about a nutrient, but they lack specific details of the quantities involved. As Hodgkins et al. (2012; p. 813) note, consumers do not need these details with summary systems as "in terms of [the product's] health utility, the decision has already been made for them."

Warnings are a directive form of FOP labeling which indicate when the product contains high levels of a nutrient that should only be consumed in a limited quantity. Warning systems have been included in strategies to reduce obesity and over-consumption of processed foods in some regions of the world. The Pan American Health Organization recommends that warnings be included on labels for food containing high levels of calories or key nutrients (saturated fats, salt, sugar). These recommendations have been implemented as mandatory in Mexico, Chile, Peru, and Uruguay (Buchanan, 2020). Warning labels from Chile are shown in Figure 1.

In the Pan American Health Organization system, labels are only required on FOP packaging when the quantity of calories or nutrient is higher than recommended. In this system there are no


FIGURE 1
Warning labels from Chile. Image taken from Grummon (2019) under creative commons attribution 4.0 international license.

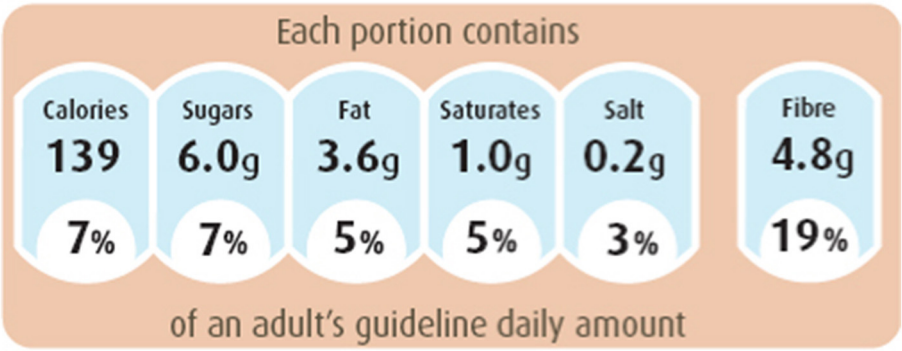



FIGURE 2
GDA label. Image from https://wiki.ead.pucv.cl/index.php/Usuario:Romina_Guerra CC BY-SA 3.0.


FIGURE 3
Traffic light labeling on food. Photograph Ian Clark/ijclark. CC BY-NC 2.0.

TABLE 1 Meaning of traffic light colors scheme colors, according to the British Nutrition Foundation (2022).

| Color | Meaning (British Nutrition Foundation, 2022) |
|---|---|
| Green | If there is mostly green on the label, then this is telling you straight away it is low in that nutrient and a healthier choice![a] |
| Amber | This means the product is neither high nor low in the specific nutrient. You can eat foods with all or mostly amber on the label most of the time |
| Red | Red does not mean you cannot eat the product, but means the food is high in fat, saturated fat, salt, or sugar. We should be cutting down on foods with lots of red on the label, or if they are eaten, to have less often and in small amounts |

[a] As pointed out by a reviewer, this explanation is slightly confusing as "mostly green" means more than one nutrient. Presumably the intention is to communicate that a green section indicates a low (and therefore healthy) level of that nutrient, and that if the label is mostly green, then the product is a healthy choice.

corresponding "low in …" labels or other indicators that a product might be a healthy option.

### 3.3. Non-directive: facts up front/reference intake/guideline daily amount

Non-directive systems include detailed information about the nutritional content of the product. However, no explicit value judgement is provided about whether the food is a healthy choice or not. As illustrated in Figure 2, the amount of each nutrient is given per portion (or per 100 g) and the label also shows the percentage that this represents of an adult's guideline daily amount. For this reason, these systems are sometimes referred to as GDA labeling or RI (reference intake).

In the United States of America, this system is referred to as "Facts Up Front" (Consumer Brands Association FMI, 2022). It displays the nutrient amount per serving both in grams/milligrams and as a percentage of a daily value (DV). These are categorized as non-directive, as they provide no indication of whether the product is a healthy choice or not.

### 3.4. Semi-directive: traffic light systems

Finally, semi-directive systems "contain information on nutrient content but also communicate decisions on healthfulness" (Hodgkins et al., 2012; p. 814). This is often achieved by Facts Up Front style labels with added color-coding, as seen in Figure 3. The most common systems use a traffic light red-green-amber distinction. As each nutrient is coded separately, these labels are sometimes referred to as Multiple Traffic Lights or MTL. In some semi-directive schemes, each nutrient is labeled as "high," "medium," or "low"" as well as, or instead of, the color-coding.

The traffic light labeling scheme is the government recommended format in the UK, and Table 1 shows the British Nutrition Foundation (2022) explanation of the coding.

As Hodgkins et al. (2012) discuss, for most food products, there will be a mixture of red, green and/or amber across the different nutrient categories. It is unusual for a product to be all red or all green. Therefore, the direction given to consumers is not as binary and clear as with the directive systems. The consumers must make a decision based on a particular nutrient or on the overall traffic light profile. For this reason, Hodgins et al. suggest that traffic light systems be classed as semi-directive and that therefore a three-way categorization labeling system is necessary. Having outlined these three categories of labels, I move on, in the next section, to give an overview of research into the effectiveness of the different systems.

## 4. Effectiveness of the labeling systems: empirical evidence

Various studies and experiments have sought to identify the most efficient way to communicate nutritional information to consumers and to thereby alter behavior in favor of more healthful food and drink choices. The discussions here draw on three systematic reviews of work in this area (Hawley et al., 2013; Hersey et al., 2013; Temple, 2020), and from these some clear patterns emerge. I then discuss several individual studies to provide more detail on the methods used and to illustrate the findings that underpin the patterns and conclusions.

The first key finding to note is that semi-directive systems appear to be more effective than the non-directive messaging. A systematic review by Hawley et al. (2013) of research into the effectiveness of food labeling found that "the MTL [multiple traffic light] label has the most consistent support" (p. 437) in terms of being beneficial to consumers. Hersey et al. (2013) similarly found that "consumers can more easily interpret nutrition information using FOP schemes that incorporate text and color to indicate "high," "medium," or "low" levels of nutrients compared with FOP labels that only display numeric information including %GDA and/or grams" (p. 12). Both reviews conclude therefore, that the semi-directive traffic light style systems are more effective than the non-directive Facts Up Front style systems. However, these reviews were carried out before the introduction of warning labels in countries such as Chile, and so they do not include directive systems in their comparisons. Temple (2020) conducted a literature search on studies published after 2011 to fill that gap and his review covers only studies that were not included in the two previous reviews (Hawley et al., 2013; Hersey et al., 2013). Although Temple notes a high level of inconsistency across the studies in his review, he concludes that the "designs for FOP labels that appear to be most successful are MTL, warning labels and Nutri-Score." Meanwhile, labels "based on GDA … were much less successful" (Temple, 2020, 5). Given these general patterns, we can look more closely at the findings of individual studies to explore the effectiveness of the labeling in more detail.

Arrúa et al. (2017) compared three labeling schemes, one from each of Hodgkins et al. categories: the GDA system (non-directive), the traffic light system (semi-directive) and the Chilean warning system (directive). Participants were asked to identify if the food products displayed on a computer screen were high in sodium. That is, they were asked to identify unhealthy products based on salt content. The participants gave correct answers in an average of 95 per cent of cases with no significant difference found between the labeling systems. There was, however, a significant difference between response times across the different labeling

systems. Response times for GDA labels were significantly longer than for the traffic light system and warning labels. Warning label response times were the fastest of all.

While warning labels appear to have the most impact when it comes to identifying unhealthy options, Adasme-Berríos et al. (2022) found that their impact was limited in other ways. Their study showed "no evidence for effects on nutritional knowledge" (p. 1547) when warning labels were used. So, while they may be the most effective in terms of individual decisions, warning labels did little to educate consumers about nutrition and health more generally.

In a follow up study, Arrúa et al. (2017) asked participants to rate the perceived healthfulness of products and the frequency with which they should consume them. The stimuli were all products that were typically consumed in the region (Uruguay), but the brands used were not commercially available there. The labels were modified so that one in each set was more healthful than the others based on one key nutrient. The task was therefore to identify the healthiest option. They found that warnings and traffic light labels performed equally well when participants were asked to identify the most healthful product.

Directive and semi-directive labels were also found to be effective by van Herpen and Trijp (2011). They compared a health tick logo directive label with both traffic lights and a Facts Up Front nutrition table, and this led them to the conclusion that "the logo seems to have an advantage, both in terms of the likelihood of attending to the label and the effect on choice. The MTL label also performs well, but the nutrition table does not enhance healthy choices beyond the level when no labels are present" (p. 158). A similar result was reported by Roberto et al. (2012) who compared consumer understanding of the non-directive Facts Up Front system with the semi-directive multiple traffic light scheme. They found that when it came to judging the levels of nutrients in a product, traffic lights were "substantially more helpful" (p. 140) than Facts Up Front.

As part of their study, Machín et al. (2017) compared GDA labels with two versions of the semi-directive traffic light system. One version used the typical red-amber-green multicolored coding while the other was monochromatic. The multicolored version used red for high levels of a nutrient and green for low. The monochromatic version used black for high and white for low. The study examined participants' perceptions of healthfulness for ultra-processed products, and it compared low-, middle- and high-income participants. The results paralleled the other studies in that the semi-directive systems outperformed the non-directive system. Both the traffic light systems led the participants to rate the ultra-processed products as lower in healthfulness than the GDA system for low-income participants (p. 336). However, Machín et al. found a difference between the two traffic light systems for some products. In certain instances, where the product contained some nutrients with low levels alongside others with high levels, the monochrome labels resulted in a lower perception of healthfulness. That is, the same products were perceived to be less healthy when the nutritional information was presented in black and white than when it was displayed in color. Machín et al. suggest that this might be because the green used in the colored system for low nutrient content carries with it associations of healthfulness, whereas in the monochrome system these nutrients were presented in a more neutral white.

A study carried out by Araya et al. (2019) looked at the effects of warning labels on different categories of food. They studied purchasing behavior in Chilean supermarkets over the year-long period in which the warning label scheme was introduced. They found that the warning labels led to "a substantial reduction in purchase probabilities of labeled breakfast cereals" (p. 16). However, they found that the labels had no effect on purchasing habits related to products in the cookies and chocolate ranges.

Overall, warning labels appear to be the most effective system, particularly when it comes to identifying products that should be avoided or limited. Semi-directive systems such as the multiple traffic lights appear to be more effective than the non-directive Facts Up Front systems, and equally as effective as warnings when it comes to identifying healthy options. Finally, in terms of consumer behavior and purchasing habits, the product type also makes a difference to whether the labeling is effective or not. To understand these patterns, I next compare the interpretive processes that consumers go through when they encounter each type and category of label. Implications from this relevance-theoretic analysis then follow in Section 6.

## 5. Relevance-theoretic analysis

In this section, I use relevance-theoretic assumptions about the processing of communicative inputs to explain the patterns of relative effectiveness of food nutritional labeling systems. As discussed in Section 4, in terms of encouraging consumers to avoid unhealthy products, directive warning labels appear to be the most effective system, followed by semi-directive traffic light systems. Non-directive Facts Up Front style labeling is the least effective system in terms of communicating information about healthfulness and influencing consumer behavior.

To understand how the different label formats might be interpreted by an individual consumer, imagine Rita as a typical health-conscious shopper. Rita is likely to hold a range of assumptions about food, nutrition, health, and food choices. These might include the assumptions in (3) to (7).

(3) If a product is healthy, I want to buy it
(4) If a product is unhealthy, I do not want to buy it
(5) If a product is high in fat, it is unhealthy
(6) If a product is high in sugar, it is unhealthy
(7) If a product is low in salt, it is healthy

How does the information in the various labels interact with these assumptions to yield cognitive effects? First consider warning labels and imagine that the product is high in fat. The warning label will follow a standard format such as the one shown in Figure 1 and with text that says, "High in fat." This input can immediately interact with Rita's assumption in (5), leading her to derive the conclusion in (8).

(8) This product is unhealthy

This conclusion is a new assumption that Rita now holds, and it can combine with the assumption in (4) to lead her to the conclusion that she does not want to buy the product. The inferential path from the input on the label to Rita's conclusion is relatively direct, and the input information combines with accessible assumptions that Rita already holds. Indeed, once health-conscious customers recognize the black octagonal symbols (Figure 1) as warnings, they will hold the assumption in (9), and they need not even read the text to reach a "don't buy" conclusion.

(9) If a product has a warning label on it, it is unhealthy

Next consider the inferential processes involved in the interpretation of the traffic light system label, as illustrated in Figure 3. Rita will see the color-coded sections with the accompanying nutritional information. Imagine that the label indicating fat content is colored red and contains the text in (10).

(10) One serving contains: Fat 6.9 g. 10% of the reference intake of an average adult

Decoding the text will provide Rita with information about the nutritional content of the product. However, the color-coding also makes assumptions accessible that then combine with Rita's existing assumptions to yield cognitive effects. Assuming that Rita is following the relevance-theoretic comprehension procedure and therefore taking the path of least effort, she will test out the most accessible interpretations first, and will stop when she has an optimally relevant interpretation. The red coloring of the label is likely to make certain assumptions accessible to Rita. Red is associated with danger or hazards (Chapanis, 1994; Braun and Silver, 1995; Pravossoudovitch et al., 2014) and has been demonstrated to induce an avoidance motivation (Mehta and Zhu, 2009). Furthermore, in the context of this labeling system, red is used as part of a traffic light system, and it is set in contrast to green and amber, making associations with "stop" highly accessible in the cultural contexts in which these labels are used. When used in the context of nutritional information, these associations with danger, avoidance, and stopping are most likely to be interpreted as communicating the assumption in (11), leading Rita to draw the conclusion in (12).

(11) If the nutritional label is red, the product is unhealthy
(12) This product is unhealthy

As with the warning label, Rita can then combine this new assumption with her existing assumption in (4) to reach the conclusion that she does not want to purchase the product. Although there is further and more detailed information available *via* the text on the traffic light label, Rita does not need to read or process this. The color alone has led her to a conclusion about the food, and there is no need for her to go to the extra effort of decoding and interpreting the nutritional information.

Finally, consider the processes that Rita goes through to interpret the non-directive Facts Up Front style version of the label. The information on these labels is presented against a single color background. In the US version, this is blue across the different nutrient categories and is the same across all labels. The textual information provided is given in (13).

(13) Per serving 6.9 g Sat Fat. 10% of DV

Notice that there are no easily accessible assumptions with which the input from this label can combine. None of the assumptions that Rita holds in (3) to (7) connect with this information, and there are no easily inferable assumptions that can bridge the gap either. The color of the label provides no useful input in this case. To derive cognitive effects from the Facts Up Front labeling, Rita would need to think about what she has already eaten and what else she plans to eat that day (or the day on which she thinks she will consume the product). Even if she has access to this information, it will be much less accessible than the more general assumptions in (3) to (7). Assuming that she persists with her interpretation of the label and works out how much else she will consume (rather than abandoning her search for relevance), she would need to access assumptions along the lines of (14) and (15).

(14) If I have already eaten or plan to eat over 90 per cent of my daily recommended allowance of fat today, it would not be healthy for me to eat a whole portion of this product.
(15) If I have not already eaten or plan to eat over 90 per cent of my daily recommended allowance of fat today, it would be healthy for me to eat a whole portion of this product.

It is only at this point that Rita can assess whether the product is a healthy choice for her and therefore whether she will purchase it or not. There are more inferential steps involved in reaching this point *via* the Facts Up Front labeling, and the steps are more complicated and vulnerable to error. Even health-conscious Rita will be unable to derive cognitive effects from these labels unless she knows and recalls the nutritional value of what else she has eaten that day.

This comparison of the interpretative processes that Rita follows in each case sheds light on the differences in effectiveness and ease of interpretation of the three systems. Warnings and traffic lights require less processing effort than the Facts Up Front system to guide Rita to an assessment of healthfulness and therefore a purchase decision. They involve fewer inferential steps and more accessible / less complicated assumptions.

We can also apply relevance-based interpretative processes to explain the differences identified by Araya et al. (2019). Warning labels reduced the probability that a customer would buy a labeled breakfast cereal but had no effect on cookies and chocolate. To understand why the effect on these products might be different, it is useful to think about the assumptions that consumers are likely to hold about them. It is likely that most consumers will be aware that cookies and chocolate products are high in sugar, fat, and calories. That is, before they see the packaging, the customers are likely to hold the assumptions in (16) to (21).

(16) Cookies are high in fat
(17) Cookies are high in sugar
(18) Cookies are high in calories

(19) Chocolate is high in fat
(20) Chocolate is high in sugar
(21) Chocolate is high in calories

Adding warning labels to these products will, therefore, have little effect. New information is relevant only if it interacts with our assumptions to lead to a cognitive effect. In this case, however, the consumer already holds assumptions about the food products with a high degree of certainty. Therefore, the information on the label is unlikely to strengthen the assumption further. If you are already 100 per cent sure that chocolate is high in sugar, a high in sugar warning label on a chocolate bar has no relevance for you. Breakfast cereals, on the other hand, are not so widely associated with high levels of fat, sugar, and calories as confectionary is. Indeed, it is likely that many consumers consider breakfast cereals to be a healthy (or at least not an unhealthy) option. The packaging designs for cereals are often used to promote properties that are associated with health. They might, for example, state on the package that the product is a source of vitamins, fiber, or iron. This may well mean that the typical consumer holds the assumption in (22).

(22) Breakfast cereals are healthy

A health-conscious consumer who also holds the assumption in (3) ("If a product is healthy, I want to buy it") may decide to purchase cereals on this basis. The information contained in warning labels, and indeed the very presence of the warning labels themselves, will, however, contradict the assumption in (22). If the customer accepts the warning labels as a reliable source of information, she will eliminate her assumption in (22), and this in turn will lead her away from a decision to buy. It is precisely because customers either hold no assumptions about the healthfulness of cereals or may hold incorrect assumptions about this, that the warning labels can change behaviors. Warning labels are relevant in such contexts. However, when the consumer already knows the product is unhealthy, the label will lead to no cognitive effects and will therefore not be relevant.

## 6. Discussion and implications for communication design

In Section 5, I demonstrated that the relevance-theoretic pragmatic framework can be used to understand the interpretive processes consumers go through when they encounter front of pack nutritional labeling. We can understand the difference in effectiveness of the labeling schemes as related to their relevance in terms of cognitive effects and processing effort. This has implications for both labeling policy and design, and it can inform the practice of communication design more broadly.

As we saw in Section 4, warnings were more effective than the other systems when it comes to identifying unhealthy options. To be effective from a health policy perspective, nutritional labeling needs to guide a consumer to a "buy" or "don't buy" conclusion in as few inferential steps as possible. It should also rely on as few contextual assumptions as possible, and those assumptions should be highly accessible or easily inferable. The information on a label will only be relevant if it can combine with contextual assumptions to yield cognitive effects. While the directive warning labels contain less information than the traffic light or Facts Up Front systems, the information that they do contain easily combines with highly accessible assumptions. Warning labels require the lowest level of background information on health and nutrition to process, and even a consumer with little or no nutritional knowledge and with no interest in healthy eating will recognize a warning sign as marking something to be avoided or treated with caution. Similarly, the avoidance and danger associations of the red color-coding (and likewise, the healthy "go" associations of the color green) do not require an interest in or knowledge of healthy lifestyle choices to interpret. Indeed, in the case of the warning labels, it is not even necessary to read the warning text. As Arrúa et al. (2017; p. 2315) point out, "warnings appeared on the labels only when the content of the target nutrient was high." The very presence of a warning-style label is enough of an input to lead the consumer to the conclusion that the product is unhealthy.

We also saw the impact of the color-coding in the findings from Machín et al. (2017) discussed in Section 4. The use of green rather than white to indicate that a nutrient's levels are low led to a product being perceived as more healthful, despite all other information on the label being the same as the white label. Accessible assumptions about green meaning "go" or being associated with health are enough to produce a different interpretation of the product's nutritional value, and consumers will access and draw conclusions from the most accessible assumptions and associations first.

The review by Hersey et al. (2013) suggests that systems which indicate "high," "medium," or "low" for each nutrient are the easiest to interpret, whether they rely on color, text, or both to communicate this information. A study by Malam et al. (2009) also found that the labels with the highest levels of comprehension overall were those "combining text (the words high, medium, low), traffic light colors and % Guideline Daily Amount (GDA)" and those "combining text and traffic light colors." Again, we can understand this in terms of the assumptions that the consumers hold. Far more consumers will hold a general assumption such as (23) than a specific assumption such as (24).

(23) High fat foods are unhealthy
(24) Men should not consume more than 30g of fat per day and women should not consume more than 20g of fat per day (National Health Service, 2020).

This means that more people will be able to conclude whether a product is healthy or not based on the "high," "medium," and "low" labels. Anyone who does not already hold an assumption such as (24) could, presumably, stop and look up health recommendations and thereby access this information. However, the more effort that is involved, the more likely it is that the customer will abandon the search for relevance and make the purchase decision based on other criteria. While the GDA labels contain the same, and indeed more, information than the "high," "medium," or "low" labels, the information requires more processing effort (for most people) and is therefore less relevant. More information does not necessarily mean better when it comes to communicating health (or other)

information. What is key is the ease of processing for as many consumers as possible. New information is easier to process if it combines with highly accessible assumptions and the more people who hold those assumptions, the wider the reach of the message will be.

It is not, however, simply a matter of the assumptions that a consumer may or may not hold. Designers of communications need to also consider the strength with which a consumer holds an assumption and the sort of information that would convince them to strengthen or eliminate that assumption, thereby generating a cognitive effect. In a report for the Foods Standards Agency, Malam et al. (2009) found that some users who are confident in their knowledge of what is and what is not healthy may not use labels at all. While they may be health conscious, if they are already highly confident in the assumptions they hold, the information on the label will be less likely to be relevant to them. When we are 100 per cent confident about something, it is not possible to strengthen that assumption, and it is much less likely that new information will contradict and eliminate it. At the other end of the customer spectrum, those who are not interested in healthy eating tend to avoid FOP labeling, according to Malam et al., because they consider it to be 'an unwelcome attempt to control their behaviour' (4). Thus, designers of health communication policies must consider not only what information to communicate and how to communicate it, but also how to encourage consumers to trust the source of the information. We will not update our assumptions if we do not trust the source of the information or if we do not consider the source to be credible (Sperber et al., 2010).

The analysis of the nutritional labeling systems also reveals that, when creating health messaging, designers should focus on the conclusion to which they wish to guide the consumers. Effective messaging is not just about the dissemination of information, but rather about producing stimuli which will lead to the intended cognitive effects. For example, encouraging people to eat healthy foods is different to encouraging people not to eat unhealthy foods. The designers must understand what assumptions the consumers already hold and think about how their messaging will interact with those. For example, the information in warning systems can only lead a customer to a "don't buy" conclusion, as it can only combine with assumptions about what not to eat. This aligns with the overall aim of reducing consumption of ultra-processed foods identified by the Pan American Health Organization and so will be an effective strategy to achieve this outcome. However, warnings are less likely to improve consumer's nutritional understanding or guide them to alternatives which are positively healthy as they contain no information which can combine with assumptions about healthful food or nutrition. The reverse is true of health endorsement directive labeling such as health tick logos (van Herpen and Trijp, 2011). These can only combine with existing assumptions about what is a healthy choice, and so while they are effective if the aim is to increase the consumption of healthy products, they have less direct impact if the aim is to decrease the purchase of unhealthy products.

In the discussions here, I have assumed that those designing the labeling want to encourage the consumption of healthy food and discourage the consumption of unhealthy foods. However, food producers may, of course, have other motivations. By understanding the interpretive stages involved in processing nutritional messaging we can also understand how it might be circumvented. For example, one way for producers of less healthy products to maintain the appearance of caring about their customers while avoiding a loss in revenue is to comply with good practice guidelines, but to present information in the least accessible, least relevant way. Therefore, the implications and lessons outlined here are intended for regulatory bodies and policy makers just as much as they are for the food producers and packaging designers.

## 7. Concluding remarks

Relevance theory as a pragmatic framework for understanding how we interpret utterances and other ostensive acts of communication provides us with a model to analyze the consumer's journey as they process a piece of messaging. We can compare different versions of the same message and link their effectiveness to the ease of interpretation for the intended audience. Communication is a cognitive process in which new information interacts with assumptions to yield effects. To communicate effectively we must consider the assumptions that the intended audience already hold, and we must be clear about the assumptions that we want them to hold. Designing effective messaging is a matter of getting the audience from one set of assumptions to the other in as few interpretative steps as possible. While I have focused on nutritional food labeling in these discussions, the approach and analyses exemplified here can be applied to other communicative contexts and has wide-reaching implications for communication design and policy more generally.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

## 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.

## Publisher's note

# References


Adasme-Berríos, C., Aliaga-Ortega, L., Schnettler, B., Parada, M., Andaur, Y., et al. (2022). Effect of warning labels on consumer motivation and intention to avoid consuming processed foods. *Nutrients* 14, 1547. doi: 10.3390/nu14081547

Araya, S., Elberg, A., Noton, C., and Schwartz, D. (2019). Identifying food labeling effects on consumer behavior, Available online at: https://ssrn.com/abstract=3195500 (accessed December 9, 2022).

Arrúa, A., Machín, L., Curutchet, M. R., Martínez, J., Antúnez, L., Alcaire, F., et al. (2017). Warnings as a directive front-of-pack nutrition labelling scheme: comparison with the guideline daily amount and traffic-light systems. *Public Health Nutr.* 20, 2308–2317. doi: 10.1017/S1368980017000866

Braun, C. C., and Silver, N. C. (1995). Interaction of signal word and colour on warning labels: differences in perceived hazard and behavioural compliance. *Ergonomics* 38, 2207–2220. doi: 10.1080/00140139508925263

British Nutrition Foundation (2022). Looking at Labels. Available online at: https://www.nutrition.org.uk/putting-it-into-practice/food-labelling/looking-at-labels/ (accessed December 9, 2022).

Buchanan, S. (2020). New warning labels now required on packaged food. Available online at: https://mexiconewsdaily.com/news/new-warning-labels-now-required-on-packaged-foods/ (accessed December 9, 2022).

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Chapanis, A. (1994). Hazards associated with three signal words and four colours on warning signs. *Ergonomics* 37, 265–275. doi: 10.1080/00140139408963644

Consumer Brands Association and FMI (2022). Facts Up Front. Available online at: http://www.factsupfront.org/ (accessed December 9, 2022).

European Union (2/11/22). Nutrition declaration - EU labelling rules - Your Europe. Available online at: https://europa.eu/youreurope/business/product-requirements/food-labelling/nutrition-declaration/ (accessed December 9, 2022).

Forceville, C. (2014). "Relevance Theory as model for analysing visual and multimodal communication," in *Visual Communication*, ed. D. Machin (Berlin: De Gruyter Mouton) 51–70.

Forceville, C. (2020). *Visual and Multimodal Communication: Applying the Relevance Principle*. Oxford: Oxford University Press.

Grummon, A. (2019). Think Before You Drink - Can Warning Labels on Sugary Beverages Raise Consumer Awareness and Encourage healthier Choices? Available online at: https://weighinginblog.org/blog/think-before-you-drink-can-warning-labels-on-sugary-beverages-raise-consumer-awareness-and-encourage-healthier-choices (accessed December 7, 2022).

Guiding Stars Licensing Company. Guiding Stars. Available online at: https://guidingstars.com/ (accessed December 9, 2022).

Hawley, K. L., Roberto, C. A., Bragg, M. A., Liu, P. J., Schwartz, M. B., and Brownell, K. D. (2013). The science on front-of-package food labels. *Public Health Nutr.* 16, 430–439. doi: 10.1017/S1368980012000754

Hersey, J. C., Wohlgenant, K. C., Arsenault, J. E., Kosa, K. M., and Muth, M. K. (2013). Effects of front-of-package and shelf nutrition labeling systems on consumers. *Nutr. Rev.* 71, 1–14. doi: 10.1111/nure.12000

Hodgkins, C., Barnett, J., Wasowicz-Kirylo, G., Stysko-Kunkowska, M., Gulcam, Y., Kustepeli, Y., et al. (2012). Understanding how consumers categorise nutritional labels: a consumer derived typology for front-of-pack nutrition labelling. *Appetite* 59, 806–817. doi: 10.1016/j.appet.2012.08.014

Hornik, R. C. (ed.) (2002). *Public Health Communication: Evidence for Behavior Change*. New Jersey: Lawrence Erlbaum Associates.

Machín, L., Cabrera, M., Curutchet, M. R., Martínez, J., Gimémez, A., and Ares, G. (2017). Consumer perception of the healthfulness of ultra-processed products featuring different front-of-pack nutrition labeling schemes. *J. Nutri. Edu. Behav.* 49, 330–338. doi: 10.1016/j.jneb.12003

Malam, S., Clegg, S., Kirwan, S., McGinigal, S., Raats, M. M., Shepherd, R., et al. (2009). Comprehension and Use of UK Nutrition Signpost Labelling Schemes. London: Food Standards Agency. Available online at: https://openresearch.surrey.ac.uk/esploro/outputs/book/Comprehension-and-use-of-UK-nutrition-signpost-labelling-schemes/99516776302346 (accessed March 27, 2022).

Mehta, R., and Zhu, R. (2009). Blue or red? Exploring the effect of color on cognitive task performances. *Science* 323, 1226–1229. doi: 10.1126/science.1169144

National Health Service (2020). Fat: The facts. Available at: https://www.nhs.uk/live-well/eat-well/food-types/different-fats-nutrition/ (accessed December 16, 2022).

Pravossoudovitch, K., Cury, F., Young, S. G., and Elliot, A. J. (2014). Is red the colour of danger? Testing an implicit red–danger association. *Ergonomics* 57, 503–510. doi: 10.1080/00140139.2014.889220

Roberto, C. A., Bragg, M. A., Schwartz, M. B., Seamans, M. J., Musicus, A., Novak, N. et al. (2012). Facts up front versus traffic light labels: a randomized controlled trial. *Am. J. Prev. Med.* 43, 134–141. doi: 10.1016/j.amepre.2012.04.022

Sasamoto, R., and O'Hagan, M. (2020). "Relevance, style and multimodality: typographical features as stylistic devices," in *Relevance Theory, Figuration and Continuity in Pragmatics*, ed. A. Piskorska (Amsterdam: John Benjamins) 193–226.

Sasamoto, R., O'Hagan, M., and Doherty, S. (2017). Telop, affect, and media design: a multimodal analysis of Japanese TV programs. *Telev. New Media* 18, 427–440. doi: 10.1177/1527476416677099

Scott, K., and Jackson, R. (2020). "When everything stands out, nothing does: typography, expectations and procedures", in *Relevance Theory, Figuration and Continuity in Pragmatics*, ed. A. Piskorska (Amsterdam: John Benjamins) p. 167–192.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., et al. (2010). Epistemic vigilance. *Mind Lang.* 25, 359–393. doi: 10.1111/j.1468-0017.2010.01394.x

Sperber, D., and Wilson, D. (1995). *Relevance: Communication and Cognition. 2nd edn*. Oxford: Blackwell.

Temple, N. J. (2020). Front-of-package food labels: a narrative review. *Appetite* 144, 104485. doi: 10.1016/j.appet.2019.104485

Tseronis, A. (2018). Determining the commitments of image-makers in arguments with multimodal allusions in the front covers of The Economist. *Int. Rev. Pragm.* 10, 243–269. doi: 10.1163/18773109-01002006

van Herpen, E., and Trijp, H. C. M. (2011). Front-of-pack nutrition labels. Their effect on attention and choices when consumers have varying goals and time constraints. *Appetite* 57, 148–160. doi: 10.1016/j.appet.04, 011.

Wakefield, M. A., Loken, B., and Hornik, R. C. (2010). Use of mass media campaigns to change health behaviour. *The Lancet* 376, 1261–1271. doi: 10.1016/S0140-6736(10)60809-4

Wilson, D., and Sperber, D. (2006). "Relevance Theory," in *The Handbook of Pragmatics*, eds. L. Horn and G. Ward (Oxford: Blackwell) p. 606–632.

Wilson, D., and Sperber, D. (2012). *Meaning and Relevance*. Cambridge: Cambridge University Press.

frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 06 July 2023
DOI 10.3389/fcomm.2023.1176845



OPEN ACCESS

EDITED BY
Stavros Assimakopoulos,
University of Malta, Malta

REVIEWED BY
Agnieszka Piskorska,
University of Warsaw, Poland
Ryoko Sasamoto,
Dublin City University, Ireland

*CORRESPONDENCE
Didier Maillat
didier.maillat@unifr.ch

RECEIVED 28 February 2023
ACCEPTED 09 June 2023
PUBLISHED 06 July 2023

CITATION
Boulat K and Maillat D (2023) Strength is relevant: experimental evidence of strength as a marker of commitment.
*Front. Commun.* 8:1176845.
doi: 10.3389/fcomm.2023.1176845



# Strength is relevant: experimental evidence of strength as a marker of commitment

Kira Boulat and Didier Maillat*

Department of English, Université de Fribourg, Fribourg, Switzerland

When relevance theory tried to express the underlying processes involved during interpretation, Sperber and Wilson posited a process of context elaboration in which interpretation is seen as a path of least effort leading to the selection of a set of most salient contextual assumptions and implications. In this view, contextual assumptions are not randomly scattered in the hearer's cognitive environment during this context elaboration process. Instead, Relevance Theory claims that there are some organizing principles ordering contextual assumptions and determining which assumptions will be more likely to be accessed first in the process. The focus of this paper is on one such organizing principle captured by the notion of strength. Sperber and Wilson define it as the degree of confidence with which an assumption is held. While this notion has been posited right from the early days of Relevance theory, it has been left relatively untouched in relevance-theoretic accounts. In this paper, we will assess the explanatory potential of the notion of strength by linking it to the much-debated range of phenomena understood as related to commitment, i.e., the degree of speaker involvement in the truth of their utterance. Our goal will be to argue for a theoretical account of strength, in which strength is regarded as a cognitive marker of commitment, and more generally of the epistemic value of an utterance. In order to support this claim, we will present a series of original experimental designs in which we manipulated the level of speaker commitment in the information conveyed by their utterance. We predicted, on the basis of the theoretical model put forward, that such a manipulation would impact the level of strength. This cognitive effect, it is claimed, can in turn be measured through a recall task. We present results which support this model and discuss its implications.

KEYWORDS

**commitment, relevance theory, strength, experimental pragmatics, certainty, epistemic vigilance, evidentiality**

## 1. Introduction

Commitment has attracted a lot of attention as it touches upon a range of central semantic and pragmatic phenomena such as truth, reported speech, modality and evidentiality, among others.[1] As such, commitment appears in the work by scholars from different linguistic fields, including the French théorie de l'énonciation, Linguistic

1 During the elaboration of the experimental studies presented here, the first author discussed the design extensively with Napoleon Katsos. We wish to acknowledge the rich feedback and insights provided in the welcoming atmosphere of his lab. We would also want to thank two reviewers for providing us with very constructive comments. The usual disclaimers remain.

Polyphony, Speech Act Theory, Argumentation Theory, and Cognitive Pragmatics. All of these approaches pointed out that a speaker cannot always be said to be held responsible for what she communicates.[2] Indeed, her degree of commitment—her level of endorsement of the information conveyed in her utterance—may vary and it can be linguistically modulated.

The purpose of this contribution is 2-fold: first it tries to further our understanding of the pragmatics of commitment phenomena by linking commitment to the properties which determine the salience of a given contextual assumption in the cognitive environment of a hearer. Thus, it offers a cognitive pragmatic model to account for the kind of processes at work when a hearer interprets an utterance and, crucially, when he has to assess the level of commitment associated with it. This model brings together the insights of the previously mentioned approaches to put forward a fine-grained, empirically testable pragmatic account of commitment. Second, this paper seeks to offer empirical evidence for the purported model by reporting on an experimental design that tests some of its most central predictions.

In Section 2, we offer some landmarks by providing a brief overview of the various approaches which have used the concept of commitment. We then proceed to propose a revised model for the analysis of commitment phenomena in a relevance-theoretic framework. In doing so, we also offer a detailed typology of commitment phenomena which allows us to identify more precisely the focus of this paper as the processes linked to the hearer's interpretation of the speaker's commitment. Section 4 develops the pragmatic model of commitment further and argues that the interpretation of the speaker's commitment to a given utterance contributes to determining the relative manifestness of the assumption derived from it in the cognitive environment of the hearer. Specifically, we claim that the perceived degree of speaker commitment to an utterance will directly influence the strength of the corresponding assumption in the hearer's cognitive environment. Based on these theoretical claims, the second part of the paper presents two experimental studies which test this main hypothesis. We conclude by discussing the results which provide support for the argument that commitment markers in a speaker's utterance have a cognitive effect on the manifestness of the corresponding assumption in the cognitive environment of the hearer.

## 2. Commitment

If commitment has long been recognized as a key aspect of communication, it has often been studied from an indirect theoretical perspective as a notion associated with some other linguistic phenomenon (see Coltier et al., 2009; Dendale and Coltier, 2011; Boulat, 2018 for an overview). Thus, even though the notion of commitment is repeatedly mentioned in contemporary linguistics, it is often combined with notions such as source, enunciation, truth, modality and assertion, to name just a few (Coltier et al., 2009; p. 7). Furthermore, scholars disagree on a number of properties associated with commitment: (a) its scope; (b) the person who is supposed to commit; (c) the type of content which one can be committed to; (d) the possibility of not being committed at all; and (e) the idea that commitment is a continuum rather than a categorical notion.

More specifically, commitment has been studied, often obliquely, through the lenses of linguistic domains such as Enunciation Theory (Culioli, 1971), Linguistic Polyphony (Ducrot, 1984; Nølke, 2001; Nølke et al., 2004; Birkelund et al., 2009), Speech Act Theory (Austin, 1975; Searle, 1979; Katriel and Dascal, 1989; Falkenberg, 1990. Argumentation Theory (Hamblin, 1970; Walton, 1992, 1993, 1996, 1997, 2008a,b; Beyssade and Marandin, 2009; Semantics (Papafragou, 2000a,b, 2006), as well as relevance-theoretic pragmatics (Sperber and Wilson, 1987/1995; Ifantidou, 2001; De Saussure, 2009; Moeschler, 2013; Vullioud et al., 2017; Mazzarella et al., 2018; Bonalumi et al., 2020).

Within the enunciative and polyphonic frameworks, commitment (referred to as *endorsement*)[3] marks the speaker's subjectivity in the utterance and encompasses a range of linguistic phenomena (such as speech acts, modality, evidentiality, reported speech, amongst others) which give rise to a complex interplay between the speaker and the utterance itself. In Speech Act Theory, the speaker is not only construed as being committed to the meaning conveyed by the utterance, but also to what is being communicated, i.e., the action that she is trying to perform when uttering that utterance. Argumentation Theory construes commitment as a property that transfers from one statement to another. From this perspective, commitment forms a set of claims (the commitment store) that an arguer can be regarded as upholding in an argumentative exchange. Commitment is therefore thought of as a mental representation that captures an argumentative standpoint. Finally, Relevance Theory has addressed commitment from different perspectives by focussing on the way commitment expressed by the speaker interacts with the comprehension procedure in the epistemic evaluation of information. For example, studies on epistemic vigilance (Mascaro and Sperber, 2009; Sperber et al., 2010; Mercier and Sperber, 2017) have shown how the epistemic vigilance mechanisms will distinguish between the degree of commitment assumed by the speaker toward the content of the utterance and her degree of commitment as a function of her reliability as a competent source for the information conveyed by that utterance.[4] In more recent approaches, scholars have investigated the impact that meaning-relations (explicit, implicit or presupposed) have on the perceived level of commitment to which a speaker can be held accountable (see Vullioud et al., 2017; Mazzarella et al., 2018; Bonalumi et al., 2020).

If each approach has attempted to find how best to represent the speaker's decision to endorse a given utterance at various degrees or to dissociate herself from that utterance, a survey of verbal aspects of commitment (see Boulat, 2018) shows that definitions and accounts of commitment markers in linguistics display the

2 In this contribution we will refer to a female speaker, whereas the hearer will be assumed to be male.

3 "Prise en charge", in French.

4 The degree of speaker commitment is only one of several dimensions that the Epistemic Vigilance filter controls for. Speaker benevolence, or informational coherence would also enter the evaluation process for instance.

same heterogeneity observed on a conceptual level. Yet linguistic markers of commitment (such as plain assertions, epistemic modals and evidential expressions) have long been identified and studied in various areas of linguistic enquiry (see Ifantidou, 2001).

More recently, pragmatic approaches to commitment have tried to capture the cognitive aspects of commitment, as they have been highlighted by certain relevance-theoretic approaches (see De Saussure, 2008, 2009; Morency et al., 2008; Vullioud et al., 2017; Mazzarella et al., 2018; Bonalumi et al., 2020) for instance. This last type of approach on commitment phenomena looks promising and crucially lends itself to experimental testing.

In what follows, we are trying to revisit the notion of commitment to propose a new take on (a) its cognitive nature and (b) the part played by graded commitment markers in triggering commitment assignment processes. In doing so, we argue for a cognitively grounded pragmatic model which captures commitment as a determining factor for the strength of contextual assumptions stored in the cognitive environment of the hearer (see Sperber and Wilson, 1987/1995).

## 3. Revisiting a pragmatic model of commitment

In this paper, we want to extend the existing pragmatic account of commitment (see Boulat, 2015, 2018; Boulat and Maillat, 2017, 2018) as it has been set within cognitive, relevance-theoretic pragmatics (Sperber and Wilson, 1987/1995) and epistemic vigilance studies (Mascaro and Sperber, 2009; Sperber et al., 2010; Mercier and Sperber, 2017). In this model, we argue that commitment accounts tend to conflate different types of commitment phenomena which need to be distinguished and that it cannot be limited to the speaker's propositional attitude and to the result of a higher-level inference on the illocutionary force. Therefore, we propose a commitment typology, which includes and refines some of the categories identified by De Saussure (2008, 2009), Morency et al. (2008), Moeschler (2013).

If scholars generally focus on a speaker-based pragmatic model of commitment, we think it is equally important to distinguish the hearer's perspective and therefore to include both utterance production and utterance comprehension phenomena in our account of commitment. Therefore, our proposal for a typology of commitment (Boulat, 2015, 2018; Boulat and Maillat, 2017, 2018), is inspired by an existing contrast in the theoretical literature on commitment between a linguistic and a cognitive focus on the one hand, and a production and comprehension focus on the other. This typology proposes to differentiate four types of commitment-related processes during a verbal interaction: speaker commitment, communicated commitment, attributed commitment and hearer commitment. The unfolding of these four different processes is illustrated in Figure 1.

In order to illustrate these typological distinctions, let us imagine a conversation between Elizabeth and Fitzwilliam, starting with utterance (1):

(1) *Elizabeth*: Jane is not a gold digger.

Before uttering (1), we must assume that Elizabeth has access to a mental representation of the assumption *Jane is not a gold digger* in her cognitive environment. This inscrutable side of commitment is what we refer to as speaker commitment, i.e., the degree of epistemic endorsement assumed by the speaker toward assumptions which are manifest in her cognitive environment.[5] In relevance-theory this epistemic property which applies to the way contextual assumptions are represented in somebody's cognitive environment is captured under the notion of *strength*, which constitutes one of two properties of assumptions which determine their degree of manifestness in the cognitive environment.

> "Manifestness depends on two factors [...]: strength of belief and salience. These factors are quite different—one is epistemic and the other cognitive—and for some purposes it would be unsound to lump them together. However, we need to consider their joint effect in order to explain or predict the causal role of a piece of information in the mental processes of an individual (Sperber and Wilson, 2015; p. 133).[6]"

Going back to speaker commitment, Sperber and Wilson (1987/1995; p. 77) explain the type of parameters which affect the strength of a given assumption in her cognitive environment. They suggest that

> [T]he initial strength of an assumption may depend on the way it is acquired. For instance, assumptions based on a clear perceptual experience tend to be very strong; assumptions based on the acceptance of somebody's word have a strength commensurate with one's confidence in the speaker; the strength of assumptions arrived at by deduction depends on the strength of the premises from which they were derived. Thereafter, it could be that the strength of an assumption is increased every time that assumption helps in processing some new information, and is reduced every time it makes the processing of some new information more difficult.

Hence, Elizabeth's assumption about her sister in example (1) is entertained with a high degree of strength. Since Elizabeth is a cooperative speaker (i.e., she wants to improve Fitzwilliam's representation of the world by giving him the opportunity to integrate an accurate piece of information in his cognitive environment), she produces utterance (1), which conveys a high degree of certainty, as it is presented as a plain assertion. Communicated commitment is thus defined as the public expression of what the speaker wants to convey about her level of commitment. Put differently, it refers to the speaker's ways of presenting her utterance with more or less certainty, and of presenting herself as more or less reliable.[7] Obviously, speakers are not always cooperative so speaker commitment and communicated commitment are not necessarily aligned.

5 As we will see later on and following up on the ideas put forward by Sperber et al. (2010) our model takes strength to be a function of the certainty of the communicated content and of the reliability of its source.

6 In a footnote linked to this discussion, Sperber and Wilson (2015) explain that the notion of salience mentioned here is equivalent to that of 'accessibility' which is used extensively in relevance-theoretic accounts.

7 Obviously, a speaker can also report some other locutor's speech, in which case it is the latter person's reliability that will be factored in when determining the degree of commitment.

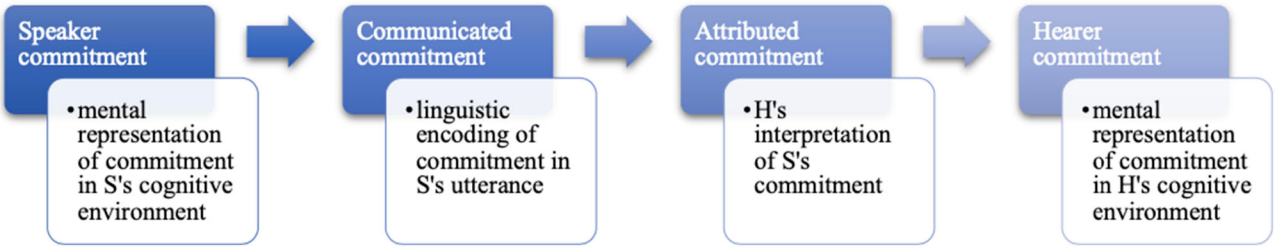



FIGURE 1
Commitment-related processes.

On the hearer's side, Fitzwilliam's understands Elizabeth's utterance and assesses the certainty of the content and Elizabeth's reliability. This is what we propose to call attributed commitment, which refers to the result of the hearer's assessment of the certainty of the communicated information and of the speaker's reliability, based on available linguistic cues and contextual assumptions. Elizabeth conveyed a plain assertion, therefore hinting at high certainty. Furthermore, Fitzwilliam knows Elizabeth well, he holds her in high esteem and thinks she is reliable. Based on this assessment of the content and of its source, Fitzwilliam integrates the piece of information *Jane is not a gold digger* in his cognitive environment. Since degrees of certainty and reliability translate into cognitive strength in our model (see below for a detailed discussion), Fitzwilliam assigns the assumption *Jane is not a gold digger* a high degree of strength. This is hearer commitment, which corresponds to the degree of strength assigned to the same piece of information as it is integrated in the hearer's cognitive environment.

Our proposition for a typology of the notion of commitment can be summarized as follows:

a. **Speaker Commitment** is the degree of strength assigned to the assumptions in the speaker's cognitive environment.
b. **Communicated Commitment** refers to the speaker's ways of explicitly presenting the piece of information with more or less certainty and reliability through the use of appropriate markers.
c. **Attributed Commitment** corresponds to the hearer's assessment of the certainty and reliability communicated by the speaker's utterance, based on available linguistic cues and contextual assumptions.
d. **Hearer Commitment** refers to the degree of strength assigned to this same piece of information as it gets integrated in the hearer's cognitive environment.

Not only does this typology distinguish the speaker's and hearer's perspective, it also draws a line between production and comprehension processes as well as the cognitive and linguistic component of commitment. Indeed, it deals with both mental representation (capturing commitment as a property of a cognitive representation) and linguistic marking (capturing commitment as a property of a linguistic form).

In the experimental design presented in this paper we explore the relationship between communicated commitment and hearer commitment showing how the linguistic markers of commitment in the speaker's utterance affects the integration of the information it conveys in the hearer's cognitive environment. In the next section, we extend our presentation of a cognitive pragmatics of commitment with a discussion of the concept of strength.

## 4. Measuring the strength of assumptions

Our alternative model of commitment is crucially built on the notion of strength which is a property of assumptions in the cognitive environment. According to Clark (2013; p. 114), most of our assumptions are tentatively entertained to varying degrees. This is what Sperber and Wilson (1987/1995; p. 75) refer to as the strength of an assumption, defined as the confidence with which it is held and as the result of its processing history (Sperber and Wilson, 1987; p. 701). Relevance Theory applies the concept of strength to all assumptions in an individual's cognitive environment. According to Ifantidou (2000; p. 139), degrees of strength are directly related to degrees of commitment. She writes that "the strength of an assumption for an individual is equated, roughly, with his degree of confidence in it" (Ifantidou, 2001; p. 73). Indeed, if the speaker chooses to use an evidential marker in her utterance, this marker is considered to affect the strength of her communicated assumptions, and therefore, her degree of commitment to the proposition expressed. In line with these authors, we claim that commitment has a bearing on the degree of manifestness of a given assumption as it influences its strength in the cognitive environment.

We argue further, in line with the claims made in Sperber et al. (2010), that the perceived commitment of the source to the information conveyed by her utterance will be determined by two factors: the degree of certainty with which the content is being communicated (by means of evidentiality markers) and the reliability of the source of information (evaluated in terms of competence and benevolence). These two notions are similar to those found in Mazzarella (2013). When she describes the mechanisms of Epistemic Vigilance posited by Sperber et al. (2010), she refers to an "alertness to the reliability of the source of information and to the believability of its content [...]."

Crucially though, in this pragmatic account of commitment, the effect that commitment has on the hearer's processing of an utterance is not evaluated in terms of an inference drawn about the credibility of the speaker, or an inference about the impact the utterance has on the social reputation of the speaker, as proposed

in recent relevance-theoretic studies of commitment phenomena (e.g., Vullioud et al., 2017; Mazzarella et al., 2018; Bonalumi et al., 2020). Instead, we want to propose that degrees of commitment leave a trace in the cognitive environment of the hearer by modifying the degree of manifestness of the representation of that utterance. Below, we consider how different linguistic markers of commitment can affect certainty and reliability, thereby altering the strength of the assumption conveyed by an utterance.

On the one hand, the kind of certainty envisaged here concerns the content of an utterance. It typically refers to the speaker's communicated assessment of the epistemic status of the state of affairs. This content can be said to be more or less certain as the speaker has the possibility to linguistically express more or less certainty via different markers, such as plain assertions, epistemic modals and evidential expressions (Papafragou, 2000a,b; Ifantidou, 2001; De Saussure, 2011; Hart, 2011; Marín-Arrese, 2011; Oswald, 2011; Wilson, 2012). Let us consider the following examples:

(2) Elizabeth is reading Mr Darcy's letter.

(3) Elizabeth may be reading Mr Darcy's letter.

(4) I think that Elisabeth is reading Mr Darcy's letter.

If the speaker utters a plain assertion as in (2), it conveys more certainty than if she modifies her utterance with an epistemic modal [see (3)] or with an evidential expression as in (4). Indeed, epistemic modals and evidential expressions are known to have either a weakening or strengthening function with respect to the speaker's commitment (Ifantidou, 2000, 2001). Therefore, the hearer assigns a degree of strength to the assumptions conveyed by the speaker's utterance, guided by these linguistic markers. We argue that this strength assignment impacts the hearer's integration of this same piece of information in his cognitive environment.

On the other hand, reliability is about the source of information, which includes two components: the speaker's reputation and her access to evidence. Following studies on epistemic vigilance (Mascaro and Sperber, 2009; Sperber et al., 2010; Mazzarella, 2013), the speaker's reputation is construed in terms of competence and benevolence. Competence refers to the fact that the speaker possesses genuine information, whereas benevolence corresponds to her wish to share her genuine knowledge with her audience. The speaker's access to evidence is the type of evidence she has when she communicates an utterance. Evidence is typically thought of as direct or indirect. The former type of evidence is usually considered more reliable than the latter. Indeed, an utterance based on direct evidence (i.e., evidence acquired via direct perception, as in 5) is presented as accurate and therefore more likely to be accepted by a hearer than an utterance based on indirect evidence, as in (6), (Cornillie and Delbecque, 2008; p. 39):

(5) I see that Mr Bingley is home.

(6) Reportedly, Mr Bingley is home.

From this perspective, (5) is more reliable than (6) because the speaker of (5) indicates that she has clear perceptual evidence about the fact that Mr Bingley is home. Yet, in (6), the evidence is marked as indirect, as the speaker uses the hearsay adverb *reportedly*, which suggests that she does not have direct evidence for what she communicates (Iten, 2005; p. 48).

According to our pragmatic model of commitment, commitment assignment processes take place in the relevance-theoretic comprehension procedure as theorized by Sperber and Wilson (1987/1995). It starts with the speaker producing an utterance of the type *commitment marker (p)*. As previously mentioned, commitment is construed as a function of both certainty (which applies to the content of an utterance) and reliability (which applies to the speaker's reliability). These different degrees of certainty and reliability, which are communicated by the speaker's utterance, are represented in the hearer's cognitive environment through the derivation of higher-level explicatures, defined as a type of explicature "which involves embedding the propositional form of the utterance [...] under a higher-level description such as a speech-act description, a propositional attitude description or some other comment on the embedded proposition" (Carston, 2002; p. 377). Following Ifantidou (2001), Papafragou (2006) and Moeschler (2013), we claim that these higher-level explicatures will determine the level of commitment assigned by the hearer to the assumption conveyed by the utterance (also echoing Katriel and Dascal, 1989 proposal). Therefore, through higher-level explicatures about the certainty and reliability associated with a given utterance, the degree of strength of the assumption will be modulated in the hearer's cognitive environment.

From this perspective, strength affects the degree of manifestness of all assumptions in the cognitive environment and can be regarded as the cognitive trace of commitment. We suggest further that if the hearer assumes the piece of information to be certain and the speaker to be reliable, then the corresponding assumptions in his cognitive environment will be assigned a high degree of strength and will be more made more manifest as a result.

For this model to be complete, we would want to be able to measure varying degrees of strength in the cognitive environment of the hearer. Interestingly, in their original discussion of strength, Sperber and Wilson identified a possible effect that varying degrees of strength could trigger. They write that:

> Understood in this way, the strength of an assumption is a property comparable to its accessibility. A more accessible assumption is one that is easier to recall (Sperber and Wilson, 1987/1995; p. 77).

On the basis of this claim, it would follow that if an assumption were conveyed with a higher degree of commitment, both in terms of its certainty and/or reliability, it would impact its accessibility in the cognitive environment and, as a result, it would be expected to affect the hearer's ability to recall that assumption.

## 5. Experimental study

Based on these theoretical considerations, our prediction about the expressed degree of certainty is based on the obvious fact that linguistic markers such as plain assertions, epistemic modals and evidential expressions indicate the communicated degree of certainty the speaker assigns to her utterance. Hence, the more the piece of information is linguistically presented as certain, the more likely the hearer is to attribute a strong commitment to the speaker (modulo his assessment of her reliability). He will then be likely to integrate this same piece of information is his cognitive environment with a high degree of strength. Thus we claim that *H1 high certainty markers (such as I am sure that, I know that, for sure, etc.) increase the degree of strength assigned to the communicated assumption in the cognitive environment of the hearer.*

Our main contention is that certainty markers impact on the acceptance of a piece of information in an individual's cognitive environment, i.e., that they influence hearer commitment. This is in line with theoretical claims about epistemic vigilance mechanisms, as the more certain the piece of information is, the less activated the hearer's epistemic vigilance mechanisms are, and the more likely its acceptance will be (see Moore and Davidge, 1989; Sabbagh and Baldwin, 2001; Jaswal et al., 2007; Sperber et al., 2010; Bernard et al., 2012 and Mercier et al., 2014; inter alia). Thus, Sperber et al. (2010, p. 369) write:

> "Factors affecting the acceptance or rejection of a piece of communicated information may have to do either with the source of the information—who to believe; or with its content—what to believe."

Following up on this claim, we argue further that commitment markers will modulate the acceptance of a communicated assumption by assigning more strength to an utterance presented with high certainty markers, while low certainty markers will lead to lower strength.

Consider examples (8–10):

(8) I am sure that Caroline Bingley is interested in Mr Darcy.

(9) I think Caroline Bingley is interested in Mr Darcy.

(10) I don't know if Caroline Bingley is interested in Mr Darcy.

Comparatively, the hearer will be more likely to accept and integrate (8) in his cognitive environment given the certainty conveyed by the propositional attitude marker *I am sure* than (9) which suggests considerably less certainty. In our view, example (10), on the other hand, does not convey any commitment since the speaker communicates that she is unable to endorse the information that *Caroline Bingley is interested in Mr Darcy* with a sufficient degree of certainty.

Results of several empirical studies using a recall or recognition paradigm (see, for instance, Birch and Garnsey, 1995; Mobayyen and de Almeida, 2005; Ditman et al., 2010; Fraundorf et al., 2010 and Spalek et al., 2014) indicate that some linguistic features (such as focusing constructions, pitch accent type, focus particles, verb complexity or pronouns) lead to a stronger representation of the utterance in the participants' cognitive environment, and hence to a higher accessibility in memory than other features. In line with these results and the theoretical link connecting strength, accessibility and recall (see previous section), we hypothesize that commitment markers (i.e., markers of certainty) will also affect cognitive processing in the same way. Indeed, we claim that *H2 the higher the certainty of a communicated content, the more accessible the assumption is in the hearer's cognitive environment*. It follows that, within a recognition paradigm [where accuracy rates provide evidence regarding the accessibility of the representation of the test utterances (Traxler, 2012; p. 191)], an assumption that is highly accessible will trigger higher recognition scores. Therefore, the more committed the hearer is to a given assumption, the easier it will be for him to remember the assumption.

We posit a link between the relevance-theoretic notion of strength and an individual's ability to access assumptions stored in his cognitive environment. According to our model, cognitive strength translates into accessibility in the hearer's cognitive environment.[8] Following Sperber and Wilson (1987/1995; p. 77), we argue that an assumption's assigned degree of strength, will affect its relative accessibility. We thus suggest that *H3 hearer commitment impacts upon how information is remembered by an individual*.

## 5.1. Experiment 1A about certainty

The aim of this first study was to test whether linguistic markers of certainty indicating different degrees of speaker commitment would impact how participants remember statements presented to them during a study phase. The predicted cognitive effect on memory was measured through accuracy in a recognition task taken after a distractor phase.

In order to test whether certainty markers impact on how participants recall given statements, linguistic markers were placed in three different groups: no-commitment markers (e.g., *I don't know*, *I'm not sure*, *I hope*); weak commitment markers (e.g., *I guess*, *I think*, *It seems*) and high commitment markers (e.g., *I am sure*, *I know*, *No doubt*).[9,10] The influence of the three groups of commitment markers was tested with a yes-no recognition task where participants were presented with 30 factual statements about a fictional narrative, in which statements were presented with linguistic markers expressing different commitment levels. Within this recognition paradigm, better recall was predicted for statements containing a high commitment marker than for those including a no-commitment marker. A graded structure across the three categories of linguistic markers was also expected.

### 5.1.1. Participants

Ninety Seven native English speaking Mturk workers from the United States aged 18 to 60 (48 female, 49 male) participated for monetary compensation to an online survey.[11] All workers provided written consent prior to taking the survey.[12]

8 Accessibility is defined as "the ease or difficulty with which an assumption can be retrieved (from memory) or constructed (on the basis of the clues in the stimulus currently being processed)" (Carston, 2002; p. 376).

9 The research leading to these experiments was funded by a Doc. Mobility fellowship from the Swiss National Science Foundation to the first author for the project entitled "Are you committed? A pragmatic account of commitment".

10 All the linguistic markers of certainty used in this experiment were tested and assessed by 41 native English speaking Mechanical Turk workers (from the United States) aged 18 to 61 (23 female, 18 male), in a pre-test (see Boulat, 2018).

11 Mturk is a crowdsourcing internet market which enables its users to post Human Intelligence Tasks (HITS) in exchange for money.

12 In order to take part in this experimental study, workers needed to be native English speakers, aged from 18 to 60 and to live in the United States. When these conditions were not met, workers were automatically redirected to the end of the survey.

### 5.1.2. Stimuli

We created 30 statements about a fictional narrative regarding a crime committed in Mr Black's house. These short factual statements were carefully controlled for number of words ($M$ = 6.03 words) and frequency. [13] The critical words were the last word of each statement ($n$ = 30) and were either previously studied or new. They were selected according to their length (1–2 syllables), part of speech (i.e., nouns) and frequency (50 to 600 occurrences per million words). New words were selected on the basis of the length, the part of speech, the frequency and the meaning of old words (i.e., the words previously studied). For example, for the following stimulus *Mr Black called his old mother*, the word *mother* was the "old" critical word (i.e., the word which had been previously studied before the recognition test) and the new word was *father*, which had not been studied before in that carrier sentence before the test.

### 5.1.3. Recognition test

This within-subject yes-no recognition task included 30 statements. Half of the statements were old (i.e., the exact same statements as presented to the participants in the study phase with the same linguistic marker of no-/weak/high commitment) whereas the remaining 15 statements were new (i.e., where only the critical word was modified and replaced by a word which was not presented in the study phase, but keeping the same linguistic marker of no-/weak/high commitment). Ten statements included a high commitment marker (e.g., *I know, I am sure*), 10 a weak commitment marker (e.g., *I guess, I think*) and 10 a no-commitment marker (e.g., *I don't know, I hope*). The statements were rotated through the different test conditions: the commitment levels (no-commitment, weak commitment, high commitment) and recognition (old vs. new). For example, the statement mentioned above *Mr Black called his old mother*, was rotated as follows through the different conditions in the study phase: *Nobody knows if Mr. Black called his old mother* (lists A and B), *Mr. Black probably called his old mother* (lists C and D) and *Mr. Black clearly called his old mother* (lists E and F). 15 trials were designed to prompt a positive response (i.e., "yes) and the other 15 trials a negative response (i.e., "no"). Six lists (i.e., A-F) were created by combining linguistic markers and old-new critical words using a Latin Square. As a result, there were 6 versions of the study, that is 6 lists of pseudo-randomized statements. Two sample pairs of stimuli used in the study phase and the recognition test are presented below.

**Study phase**

*Mr. Black clearly called his old mother*

*I am unsure whether the old lady found a picture*

…

**Recognition test**

*I am unsure whether the old lady found a picture (correct answer: yes)*

*Mr. Black clearly called his old father. (correct answer: no)*

…

### 5.1.4. Procedure

The experiment started with a consent form, a few demographics questions (i.e., age, gender and languages spoken) and with an on-screen instruction informing participants of the structure of the experiment. Participants were told that they would read statements the police got from a witness, about a crime committed in Mr Black's home. Participants were asked to carefully read the 30 statements provided by the witness. However, the format of the memory task was not specified. Participants were told that the to-be-recalled statements would appear briefly on the screen, for 3 s, during the study phase. After 3 practice trials, participants were warned that the task was about to start.

In line with Birch and Garnsey (1995) and Ditman et al. (2010) studies, statements were visually and individually presented for 3 s and appeared one at a time, before disappearing. Statements were presented on the screen black on white using the font Times New Roman, size 14 pt.They were then followed by the question "how would you evaluate the certainty of this piece of information?" Participants had to rate the statements on a 5-point Likert scale (where 1 = uncertain that it is the case and 5 = absolutely certain that it is the case). The ranking task was not timed so participants could answer at their own pace. The rationale for using a certainty rating as well as for not specifying the format of the memory task was to ensure that participants would process the whole statements (and not overlook the linguistic marker of certainty). Each participant was presented 30 statements and none was presented the same statement more than once. The experiment lasted 15 to 20 min. Following Ditman et al. (2010) design, a delay was placed between the study phase and the recognition test. Participants had to answer 60 simple arithmetic questions. This distractor task took ∼10 min to complete.

After answering the 60 arithmetic questions, a message appeared on the screen and informed the participants that their memory of the statements would be tested. Participants were also told that they would be presented with the question "Did the witness say the following to the police?", which would be followed by a statement such as *Mrs Lily loved dark chocolate*. The participants had to indicate whether the statement they would be presented with was one of the statements they previously read in the study phase or not. They were asked to tick the "yes" box only if the statement was exactly the same (e.g., *Mrs Lily loved dark chocolate*). However, they had to tick the "no" box if the statement was not exactly the same (for instance, if they were presented with the statement *Mrs Lily loved white chocolate*).

Participants finally took the yes-no recognition task where the last word of each statement was either old or new (e.g., *I am unsure whether the old lady found a picture/paper* or *The butler clearly moved to the North/South*). Participants were asked to answer "yes" or "no" for the 30 trials which were individually and randomly presented, in line with Ditman et al. (2010) design. When participants correctly ticked the "yes" box when the statement was

13 The words in the 30 statements were obtained from Kucera and Francis (1967) list providing the 2200 most frequent English words (see http://www.auburn.edu/~nunnath/engl6240/kucera67.html), following Birch and Garnsey (1995), Chan and McDermott (2007) as well as Haist et al. (1992) studies on memory and recognition.

TABLE 1 Fixed effects (experiment 1a).

| | Estimate | Std. error | Z-value | Pr (>\|z\|) |
|---|---|---|---|---|
| No-commitment | 0.58152 | 0.09141 | 6.362 | 1.99e-10*** |
| Weak commitment | 0.08445 | 0.09716 | 0.869 | 0.384727 |
| High commitment | 0.33379 | 0.09923 | 3.364 | 0.000769*** |

Significance codes: ***0, **0.001, *0.01, 0.05.

TABLE 2 Random effects (experiment 1a).

| | |
|---|---|
| Participants | 0.19 |
| Statements | 0.04 |

old, it was recorded as a correct answer whereas if they incorrectly pressed "yes" when it was a new statement, it was scored as an incorrect answer.

### 5.1.5. Results

We used R (R Core Team, 2016) and the *lme4* package (Bates et al., 2015) to run a generalized linear mixed effects analysis (with only random intercepts) of the interaction between commitment markers (i.e. the fixed effect) and accuracy. The analysis shows that the two categories of no-commitment and high commitment are good predictors for accuracy in the recognition task (see Tables 1, 2).

Converting the log odds given in the model (under "estimate" in Table 1) provides us with the probability of correct answers in the recognition task in each category of commitment markers: no-commitment category (0.64), weak commitment category (0.66, not significant $p = 0.38$) and high commitment category (0.71, $p = 0.0007$).

Given these results, we can say that commitment marker categories affect accuracy in the recognition task ($\chi 2\ (2) = 12.16$, $p = 0.002283$). Figure 2 below shows a slight increase in accuracy between the no-commitment and the weak commitment (labeled "med" for medium in the graph) categories. Even though there is no statistically significant difference between the two categories, the expected graded trend is visible.

### 5.1.6. Discussion

Results of experiment 1a indicate that statements containing a high commitment marker were recalled significantly better than statements containing a no-commitment marker. These results are compatible with the predictions of our pragmatic model of commitment since the cognitive impact on the processing of utterances correlates with the level of commitment expressed in the stimulus, specifically through the use of certainty markers.

However, further analyses revealed a possible interaction between the length of the statements and accuracy. Indeed, it was found that the mean of syllables per linguistic marker category might have affected the results (i.e., for no-commitment markers, $M = 4$ syllables; for weak commitment markers, $M = 3.7$ syllables, and for high commitment markers, $M = 2.5$). Since literature on recall and recognition shows that longer words or longer utterances are harder to recall than their shorter counterparts, it is possible that participants' high accuracy rate in the high-commitment condition is due to the reduced length of the statements (and not to the fact that information conveying certainty is recalled better than information conveying uncertainty).

In order to check for this eventuality, the number of syllables per statement was factored in as an independent variable.[14] Compared against the model with accuracy as the independent variable, we see that both models provide a good model fit ($p = 0.002$ when accuracy is the independent variable and $p = 0.0001$ when the length of the statement is). This could indicate a potential hidden variable in our initial model (namely, the length of the statement) which might explain the observed effect.

## 5.2. Experiment 1B

In this second study, our goal was to address some of the limitations identified in the first design and to rule out the possibility that the effect observed there was the result of a confounding factor. For that purpose, a new version of the same experimental design was set up which controlled for additional parameters.

In order to confirm that commitment markers triggered the observed effect and to discard the hypothesis that it was due to the length of the statements, we controlled length across the different types of certainty markers in experiment 1b. Longer linguistic markers were matched to shorter statements (in terms of syllables) and shorter linguistic markers were matched to longer statements (see Table 3).

Furthermore, it was also noticed that the randomization of stimuli in experiment 1a was not optimal. For instance, the first five and last five statements contained too many linguistic markers of the same category in some lists, which may have led to primacy and recency effects. As a result, particular attention was paid to the randomization of statements in experiment 1b (the lists used in the study phase are provided in the online repository).

Finally, experiment 1b addresses the potential criticism that experiment 1a might be task-specific. Since participants were explicitly asked to rate the degree of certainty of the linguistic markers after reading them in a statement, the instructions might have made participants aware of what was really being tested and this might have biased their processing of the statements. To overcome this possible criticism, the ranking task was removed from experiment 1b.

### 5.2.1. Participants

One hundred and thirty three native English-speaking Mechanical Turk workers (from the United States) aged 18 to 60 (60 female, 73 male) participated for monetary compensation. All workers provided written consent prior to taking the survey.

14 The total of number of syllables takes into account the linguistic marker and the statement (e.g., *Obviously, the old lady saw a plane* = 10 syllables).

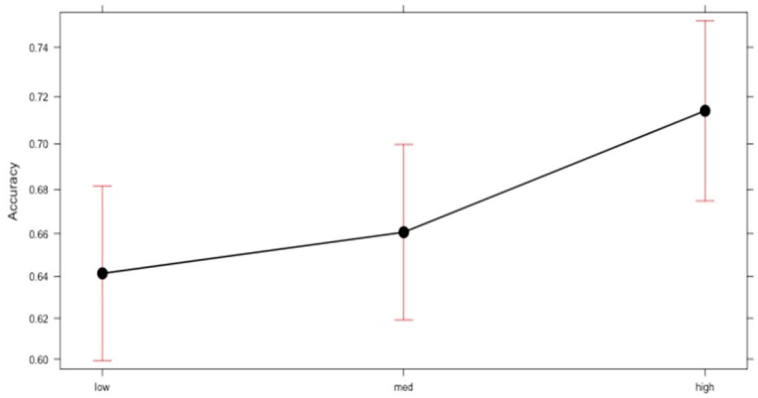



FIGURE 2
Accuracy rates by commitment category (experiment 1a).

TABLE 3 Average statement length for the 3 commitment categories.

| | List A: mean syllables | List B: mean syllables | List C: mean syllables | List D: mean syllables | List E: mean syllables | List F: mean syllables |
|---|---|---|---|---|---|---|
| No-commitment | 11.8 | 11.6 | 11.7 | 11.4 | 11.8 | 11.8 |
| Weak commitment | 11.3 | 11.5 | 11.6 | 11.6 | 11.7 | 11.5 |
| High commitment | 11.5 | 11.5 | 11.3 | 11.6 | 11.2 | 11.4 |

TABLE 4 Fixed effects (experiment 1b).

| | Estimate | Std. Error | Z value | Pr (>\|z\|) |
|---|---|---|---|---|
| No-commitment | 0.44887 | 0.07609 | 5.899 | 3.66e-09*** |
| Weak commitment | 0.01483 | 0.08434 | 0.176 | 0.8604 |
| High commitment | 0.18534 | 0.09098 | 2.037 | 0.0416* |

Significance codes: ***0, **0.001, *0.01, 0.05.

TABLE 5 Random effects (experiment 1b).

| | |
|---|---|
| Participants | 0.09 |
| Statements | 0.04 |

### 5.2.2. Materials

Thirty short factual statements were used in experiment 1b as in experiment 1a. However, the statements in the 6 lists were randomized in such a way that the first and last five statements would not display more than 2 items of the same commitment category, to avoid primacy and recency effects.

### 5.2.3. Procedure

After agreeing to participate in the survey and answering a few demographics questions, participants were told that they would read statements the police got from a witness, regarding a crime committed in Mr Black's house. They were instructed to carefully read the 30 statements provided by the witness. The format of the memory task was not specified. Participants were warned that during the study phase, the to-be-recalled statements would appear on the screen for 3 s. Then, participants performed 3 practice trials. During the study phase, each participant was presented 30 statements and nobody was presented the same statement more than once. The experiment lasted 15 to 20 min. The distractor task and the recognition test were similar to those in experiment 1a.

### 5.2.4. Results

Our model uses commitment level as a fixed effect (i.e., no-commitment, weak commitment, and high commitment, based on the pre-test results) and as a categorical predictor for accuracy in the recognition task. The analysis shows that the two categories of no- and high commitment are good predictors for accuracy in the recognition task (as shown in Tables 4, 5):

Converting the log odds given in the model (under "estimate," in Table 4) provides us with the probability of correct answers in the recognition task in each category of commitment markers: no-commitment category (0.61), weak commitment category (0.61) and high commitment category (0.65, $p < 0.05$).

In light of these results, we can say that our findings are consistent with the predicted effect that commitment marker categories should have on accuracy in the recognition task ($\chi 2$ (2) $= 5.17$, $p = 0.0754$). Results indicate that commitment markers significantly impact the accessibility of assumptions.

Figure 3 shows a slight increase in accuracy between the no-commitment and the weak commitment categories (the weak commitment category is labeled "med" for medium in the plot

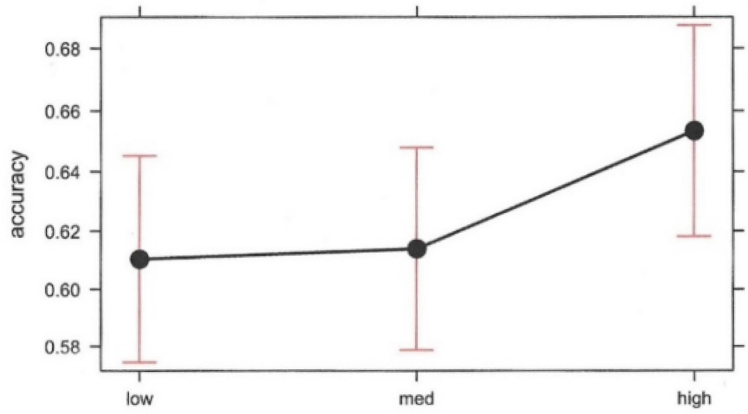



FIGURE 3
Accuracy rates by commitment category (experiment 1b).

below). Even though there is no statistically significant difference between the two categories, the expected graded trend is visible.

### 5.2.5. Discussion

The present results replicate the findings in study 1a, suggesting that participants remember statements conveying certainty differently than statements conveying uncertainty. Indeed, the participants' performance was significantly affected by commitment markers of high certainty. Once the ranking task had been removed and the stimuli controlled for length across all conditions, there is still a significant difference between the two categories of no-commitment and high commitment in terms of accuracy of recognition, even though the observed effect is weaker than in experiment 1a. Specifically, results show better retention of statements when participants were presented with a high commitment marker than with a no-commitment marker.

## 6. General discussion

Overall, our findings are fully in line with the predictions presented earlier and support our relevance-theoretic model of commitment which posits that commitment (as it is influenced by the communicated degree of certainty about the information conveyed) determines the strength of the contextual assumptions derived from the interpretation of a given utterance, which, in turn, affects the accessibility of these assumptions in a recall task.

Specifically, our findings provide supporting evidence for Hypothesis 1 which states that *high certainty markers (such as I am sure that, I know that, for sure, etc.) increase the degree of strength assigned to the communicated assumption in the cognitive environment of the hearer*. Moreover, it also goes toward confirming the relationship between communicated commitment, as expressed by the speaker in the utterance by means of linguistic markers, and hearer commitment, as measured by the strength of the assumption derived from that utterance.

Hypothesis 2 (*the higher the certainty of a communicated content, the more accessible the assumption is in the hearer's cognitive environment)* concerns the theoretically motivated relation between the relevance-theoretic notion of strength and the relative accessibility of a mental representation stored in the cognitive environment of the hearer. It appears that the early claims by Sperber and Wilson (1987/1995) about the impact of strength on recall are vindicated by these findings.

Crucially, because certainty was manipulated in these experiments as a parameter which determines the degree of commitment expressed by the speaker toward the information conveyed by her utterance, we can take our findings to speak in favor of a pragmatic model of commitment in which (communicated) commitment has a direct impact on the manifestness of an assumption. In particular, assumptions conveyed with a high level of commitment are more manifest to the hearer, than assumptions conveyed with a weaker level of commitment, as predicted in Hypothesis 3 (*hearer commitment impacts upon how information is remembered by an individual*).

Furthermore, although our statistical models are unable to tease out the intermediate commitment category (*medium*) from the other two, the expected trend can be observed between them. These promising results call for further investigation of the theoretically motivated graded structure of strength in the hearer's cognitive environment. In addition, they also call for an extension of the experimental paradigm to tap into the equally theoretically motivated effect that the source's reliability is predicted to have on strength.

To conclude, these results appear to open interesting perspectives in the study of commitment phenomena both on a theoretical level by linking commitment to manifestness in the cognitive environment; and on a methodological level by providing a new experimental design to investigate commitment assignment phenomena in pragmatics. Incidentally, they also open a new testing ground for the very central notion of strength within Relevance Theory.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Material: https://figshare.com/s/8b348e17abb03598345b, Datasets: https://figshare.com/s/d010967de38bd88d3486, R Scripts: https://figshare.com/s/f52a4b8ce48aa2852833.

## Ethics statement

The studies involving human participants were reviewed and approved by Cambridge University. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

The experiments presented in this paper were carried out as part of KB's project under the supervision of DM. All authors have contributed to the discussion, preparation, and revision of the manuscript.

## Funding

KB benefitted from a DOC.Mobility grant to study at Cambridge University for her project entitled "Are you committed? A pragmatic account of commitment" (P1FRP1_155140/2).

## 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.

## Publisher's note



## References

Austin, J. (1975). *How to Do Things with Words*. Oxford, New York: Oxford University Press.

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

Bernard, S., Hugo, M., and Fabrice, C. (2012). The power of well-connected arguments: early sensitivity to the connective *because*. *J. Exp. Child Psychol.* 111, 128–135. doi: 10.1016/j.jecp.07003

Beyssade, C., and Marandin, J. M. (2009). Commitment: attitude propositionnelle ou attitude dialogique? *La notion de Prise en Charge en Linguistique*, ed. D. Coltier, P. Dendale and Philippe de Brabanter. Langue Française 162. Paris: Editions Armand Colin, p. 89–107.

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

Birkelund, M., Henning, N., and Rita, T. (eds). (2009). *La Polyphonie Linguistique*. Langue Française 164. Paris: Editions Armand Colin.

Bonalumi, F., Scott-Phillips, T., Tacha, J., and Heintz, C. (2020). Commitment and communication: are we committed to what we mean, or what we say? *Lang. Cogn.* 12, 360–384. doi: 10.1017/langcog.2020.2

Boulat, K. (2015). "Hearer-oriented processes of strength assignment: a pragmatic model of commitment," *Evidentiality and the Semantics Pragmatics Interface*. eds B. Cornillie and J. I. Marín Arrese. [BJL 29], p. 19-39.

Boulat, K. (2018). *It's All About Strength: Testing A Pragmatic Model of Commitment*. PhD Dissertation, MS. University of Fribourg, Fribourg.

Boulat, K., and Maillat, M. (2017). She said you said I saw it with my own eyes: a pragmatic account of commitment. *Formal Models in the Study of Language: Applications in Interdisciplinary Contexts,* eds J. Blochowiak, S. Durrlemann, C. Laenzlinger, 261-279. Paris: Springer.

Boulat, K., and Maillat, M. (2018). "Be committed to your premises, or face the consequences: a pragmatic analysis of commitment inferences," in *Argumentation and Inference: Proceedings of the 2nd European Conference on Argumentation*, eds Oswald, S. and Maillat, D. Fribourg (London: College Publications) 2017, 79–92.

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell. doi: 10.1002./9780470754603

Chan, J. C.K., and McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 33, 431–437. doi: 10.1037/0278-7393.33.2.431

Clark, B. (2013). *Relevance theory*. Cambridge: Cambridge University Press. doi: 10.1017./CBO9781139034104

Coltier, D., Patrick, D., and Philippe, d. e. B. (eds). (2009). *La notion de prise en charge en linguistique*. Langue Française. (162). Paris: Armand Colin. doi: 10.3917./lf.162.0003

Cornillie, B., and Delbecque, D. (2008). Speaker commitment: back to the speaker. Evidence from Spanish alternations. *Commitment*, ed. by Philippe De Brabanter and Patrick Dendale, 37-62. Amsterdam: John Benjamins. doi: 10.1075./bjl.22.03cor

Culioli, A. (1971). Modalité. *Encyclopédie Alpha, t. 10.* Paris, Grange Batelière et Novare, Istitutogeografico de Agostini. 4031.

De Saussure, L. (2008). L'engagement comme notion cognitive associée au destinataire. *L'analisi linguistica e letteraria* 16, 475–488.

De Saussure, L. (2011). Discourse analysis, cognition and evidentials. *Disco. Stud.* 13, 781–788. doi: 10.1177/1461445611421360b

De Saussure, L. and Oswald, S. (2009). Argumentation et engagement du locuteur: Pour un point de vue subjectiviste. *Nouveaux Cahier de Linguistique Française* 29, 215–243.

Dendale, P., and Coltier, D. (2011). *La prise en charge énonciative: Etudes théoriques et empiriques.* Paris, Bruxelles: De Boeck, Duculot.

Ditman, T., Tad, T. B., Caroline, R. M., and Holly, A. T. (2010). Simulating an enactment effect: pronouns guide action simulation during narrative comprehension. *Cognition* 115, 172–178. doi: 10.1016/j.cognition.10014

Ducrot, O. (1984). *Le dire et le dit.* Paris: Editions de Minuit.

Falkenberg, G. (1990). Searle on sincerity. *Speech Acts, Meaning, and Intentions: Critical Approaches to the Philosophy* of *John R. Searle*, ed. Armin Burkhardt (Berlin, New-York: Walter de Gruyter), p. 129-146. doi: 10.1515./9783110859485.129

Fraundorf, S. H., Duane, G. W., and Aaron, S. B. (2010). (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *J. Mem. Lang.* 63, 367–386. doi: 10.1016/j.jml.06004

Haist, F., Arthur, P. S., and Larry, R. S. (1992). On the relationship between recall and recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 691–702. doi: 10.1037/0278-7393.18.4.691

Hamblin, C. L. (1970). *Fallacies*. London: Methuen.

Hart, C. (2011). Legitimizing assertions and the logico-rhetorical module: evidence and epistemic vigilance in media discourse on immigration. *Disco. Stud.* 13, 751–769. doi: 10.1177/1461445611421360

Ifantidou, E. (2000). Procedural encoding of explicatures by the Modern Greek particle *taha*. *Pragmatic Markers and Propositional Attitude*, ed A. Gisle and G. Thorstein (Amsterdam, Philadelphia: John Benjamins), p. 119-144.

Ifantidou, E. (2001). *Evidentials and Relevance*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Iten, C. (2005). *Linguistic Meaning, Truth Conditions and Relevance: The Case of Concessives*. Houndmills: Palgrave Macmillan.

Jaswal, Vikram, K., and Lauren S. M. (2007). Turning believers into skeptics: 3-years-olds' sensitivity to cues to speaker credibility. *J. Cogni. Develop.* 8, 263–283. doi: 10.1080./15248370701446392

Katriel, T., and Dascal, M. (1989). Speaker's commitment and involvement in discourse. *From Sign to Text: A Semiotic View of Communication*, ed. T. Yishai. (Amsterdam, Philadelphia: John Benjamins). p. 275-95.

Kucera, H., and Francis, N. F. (1967). Computational analysis of present day American English. Providence, RI: Brown University Press. Available online at: http://www.auburn.edu/~nunnath/engl6240/kucera67.html (accessed September 26, 2017).

Marín-Arrese, J. I. (2011). Epistemic legitimizing strategies: commitment and accountability in discourse. *Disco. Stud.* 13, 789–797. doi: 10.1177/1461445611421360c

Mascaro, O., and Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance toward deception. *Cognition* 112, 367–380. doi: 10.1016/j.cognition.05012

Mazzarella, D. (2013). "Optimal relevance" as a pragmatic criterion: the role of epistemic vigilance. *UCL Work. Pap. Linguist.* 25, 20–45.

Mazzarella, D., Robert, R., Ira, N., and Hugo, M. (2018). Saying, presupposing and implicating: how pragmatics modulates commitment. *J. Pragm.* 133, 15–27. doi: 10.1016/j.pragma.05009

Mercier, H., Bernard, S., and Clément, F. (2014). Early sensitivity to arguments: how pre-schoolers weight circular arguments. *J. Exp. Child Psychol.* 3, 11. doi: 10.1016/j.jecp.11011

Mercier, H., and Sperber, D. (2017). *The Enigma of Reason*. Cambridge, Massachusetts: Harvard University Press.

Mobayyen, F., and de Almeida, R. (2005). The influence of semantic and morphological complexity of verbs on sentence recall: implications for the nature of conceptual representation and category-specific deficits. *Brain and Cognition* 58, 168–171. doi: 10.1016/j.bandc.08039

Moeschler, J. (2013). Is a speaker-based pragmatics possible? Or how can a hearer infer a speaker's commitment? *J. Prag.* 48, 84–97. doi: 10.1016/j.pragma.11019

Moore, C., and Davidge, J. (1989). The development of mental terms: pragmatics or semantics? *J. Child Lang.* 16, 633–641. doi: 10.1017/S030500090001076X

Morency, P., Oswald, S., and de Saussure, L. (2008). Explicitness, implicitness and commitment attribution: a cognitive pragmatic approach. *Commitment*, eds B. De Philippe and D. Patrick (Amsterdam: John Benjamins). p. 197–220.

Nølke, H. (2001). La ScaPoLine 2001. Version révisée de la théorie scandinave de la polyphonie linguistique. *Polyphonie: linguistique et littéraire* 3, 44–65.

Nølke, H., Kjersti, F., and Coco, N. (2004). *ScaPoLine: La théorie scandinave de la polyphonie linguistique*. Paris: Kimé.

Oswald, S. (2011). From interpretation to consent: arguments, beliefs and meaning. *Disc. Stud.* 13, 806–814. doi: 10.1177/1461445611421360e

Papafragou, A. (2000a). *Modality: Issues in the Semantics-Pragmatics Interface*. Oxford: Elsevier Science.

Papafragou, A. (2000b). On speech-act modality. *J. Pragm.* 32, 519–538. doi: 10.1016/S0378-2166(99)00062-4

Papafragou, A. (2006). Epistemic modality and truth conditions. *Lingua* 116, 1688–1702. doi: 10.1016/j.lingua.05009

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/.

Sabbagh, M. A., and Baldwin, D. (2001). Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development. *Child Development* 72, 1054–1070. doi: 10.1111./1467-8624.00334

Searle, J. R. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Spalek, K., Nicole, G., and Isabell, W. (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.09001

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, et al. (2010). Epistemic vigilance. *Mind Lang.* 25, 359–393. doi: 10.1111/j.1468-0017.2010.01394.x

Sperber, D., and Wilson, D. (2015). Beyond Speaker's Meaning. *Croat. J. Philosophy* 15, 117–149.

Sperber, D., and Wilson, W. (1987). Précis of relevance: communication and cognition. *Behav. Brain Sci.* 10, 697–710. doi: 10.1017/S0140525X00055345

Sperber, D., and Wilson, W. (1987/1995). *Relevance: Communication and Cognition*. (2nd edition). Malden, Mass: Blackwell Publ.

Traxler, M. J. (2012). *Introduction to Psycholinguistics: Understanding Language Science*. Wiley-Blackwell: Malden, Oxford.

Vullioud, C., Clément, F., Scott-Phillips, T., and Mercier, H. (2017). Confidence as an expression of commitment: why misplaced expressions of confidence backfire. *Evol. Human Behav.* 38, 9-17. doi: 10.1016/j.evolhumbehav.06002

Walton, D. (1992). *The Place of Emotion in Argument*. University Park Pa.: The Pennsylvania State University Press.

Walton, D. (1993). Commitment, types of dialogue, and fallacies. *Inform. Logic* 14, 93–103. doi: 10.22329/il.v14i2.2532

Walton, D. (1996). *Arguments from Ignorance*. University Park Pa.: The Pennsylvania State University Press.

Walton, D. (1997). *Appeal to pity: Argumentum ad Misericordiam*. New-York: State University of New York Press.

Walton, D. (2008a). *Informal logic: A pragmatic approach*. Cambridge: Cambridge University Press.

Walton, D. (2008b). *Witness Testimony Evidence: Argumentation, Artificial Intelligence and Law*. Cambridge: Cambridge University Press.

Wilson, D. (2012). Modality and the conceptual-procedural distinction. *Relevance theory: More than understanding*, eds E. Walaszewska and A. Piskorska (Cambridge Scholars Publishing), p. 23-43.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 July 2023
DOI 10.3389/fpsyg.2023.1187343



OPEN ACCESS

EDITED BY
Kate Scott,
Kingston University, United Kingdom

REVIEWED BY
Ryoko Sasamoto,
Dublin City University, Ireland
Didier Maillat,
Université de Fribourg, Switzerland

*CORRESPONDENCE
Robyn Carston
✉ robyn.carston@ucl.ac.uk

RECEIVED 15 March 2023
ACCEPTED 15 May 2023
PUBLISHED 28 July 2023

CITATION
Carston R (2023) The relevance of words and the language/communication divide.
*Front. Psychol.* 14:1187343.
doi: 10.3389/fpsyg.2023.1187343



# The relevance of words and the language/communication divide

Robyn Carston*

Linguistics, University College London, London, United Kingdom

First, the wide applicability of the relevance-theoretic pragmatic account of how new (*ad hoc*) senses of words and new (*ad hoc*) words arise spontaneously in communication/comprehension is demonstrated. The lexical pragmatic processes of meaning modulation and metonymy are shown to apply equally to simple words, noun to verb 'conversions', and morphologically complex cases with non-compositional (atomic) meanings. Second, this pragmatic account is situated within a specific view of the cognitive architecture of language and communication, with the formal side of language, its recursive combinatorial system, argued to have different developmental, evolutionary and cognitive characteristics from the meaning side of language, which is essentially pragmatic/communicative. Words straddle the form/meaning (syntax/pragmatics) divide: on the one hand, they are phrasal structures, consisting of a root and variable numbers of functors, with no privileged status in the syntax; on the other hand, they are salient to language users as basic units of communication and are stored as such, in a communication lexicon, together with their families of related senses, which originated as cases of pragmatically derived (*ad hoc*) senses but have become established, due to their communicative efficacy and frequency of use. Third, in an attempt to find empirical evidence for the proposed linguistic form-meaning divide, two very different cases of atypical linguistic and communicative development are considered: autistic children and deaf children who develop Homesign. The morpho-syntax (the formal side of language) appears to unfold in much the same way in both cases and is often not much different from that of typically developing children, but they diverge markedly from each other in their communication/pragmatics and their development of a system (a lexicon) of meaningful words/signs.

KEYWORDS

**lexical pragmatics, relevance theory, metonymy, root-based syntax, non-compositional meaning, communicational lexicon, form/meaning divide, decoded linguistic meaning**

## 1. Introduction

Relevance theory (RT) provides a richly interdisciplinary framework for the investigation of communication and language, and it has been fruitfully employed by psychologists, philosophers, translation theorists and literary specialists, among others. It has, however, been criticized for its lack of interaction with core areas of linguistics, specifically work on linguistic structure (morphology and syntax) (Smith, 2019), a somewhat ironical situation, given that the theory typically finds its academic home in departments of linguistics. This paper, which focuses centrally on words and their meanings, continues the interaction of relevance-theoretic pragmatics with both philosophy of language and empirical psychology, while also suggesting how one aspect of its interface with the computational core of language might work.

In Section 2, I first look at the phenomenon of '*ad hoc* concepts/senses' within relevance-based lexical pragmatics and a recent application of this notion in the philosophy of language; then, I move to a related but importantly different notion of '*ad hoc* words' and the role of metonymy in their creation, ending with thoughts about the fundamental nature of the kinds of associative connections typical of metonymy, which appear to be basic and ubiquitous in our language use and in communication more generally. In Section 3, a distinction between language (construed in narrow linguistic/computational terms) and its communicative use is adopted, with words straddling the divide (having both morpho-syntactic structure and pragmatically-originated meanings). The main import of this section is to show how a syntactic treatment of words as phrasal structures, on the one hand, and the lexical pragmatic account of new word meanings, on the other, come together in explaining non-compositional word meanings (and indeed the very notion of a 'word'). The section ends with thoughts about the kind of lexicon that sits best with this pragmatically-based view of polysemy. Section 4 is devoted to presenting evidence from two profiles of atypical linguistic and/or communicative development, which, with some provisos, points to the distinct and dissociable trajectories of the formal (morpho-syntactic) and the conceptual-semantic, thus further supporting the position that this constitutes a natural divide in human cognitive architecture. I end with a short discussion of the language 'code' (syntax and lexicon), which provides rich evidential input about the speaker's meaning for the relevance-based pragmatics to work with.

## 2. Words: linguistic decoding and pragmatic inference

### 2.1. Lexical meaning adjustment and *ad hoc* concepts

From their earliest work on relevance theory, Sperber and Wilson (1986/1995) have drawn a fundamental distinction between a code model of communication and an inferential model, emphasizing that what a speaker means, what she intends her audience to grasp, when she produces a linguistic utterance is seldom, probably never, fully encoded in the linguistic meaning of the expression(s) employed. The stable established meaning provided by the linguistic components of the utterance typically (sometimes radically) underdetermines the meaning communicated. The insight comes from Grice (1967) and Donnellan (1966) in the first instance, but Grice seems to have confined pragmatic inference ('conversational logic', in his terms) to the recovery of a speaker's implicit meaning (implicatures) while viewing the explicitly communicated meaning ('what is said') as essentially encoded.[1] Since then, there has been much work in the RT framework that has demonstrated the role of pragmatic inference in contributing to the proposition explicitly communicated (termed 'explicature' in RT). This includes processes of disambiguation, saturation (e.g., assigning referents to pronouns and other indexical elements) and free enrichment (i.e., recovering components of meaning in the absence of any linguistic mandate to do so). The latter includes cases of 'unarticulated constituents' of propositional content (Carston, 2002; Recanati, 2002; Carston and Hall, 2017), but also cases where a linguistically provided meaning is pragmatically modulated so as to deliver a contextually relevant '*ad hoc*' concept/sense for the word or phrase. I focus on the latter here; that is, cases where an established sense of a word or phrase is retrieved from the memorized store (the lexicon), as part of the linguistic decoding process, but is adjusted by relevance-based pragmatic inference.[2]

According to the RT lexical pragmatics account, the forming of an *ad hoc* occasion-specific meaning or sense for a word is a consequence of standard pragmatic processes of selecting contextual assumptions, drawing cognitive implications from the utterance in this context and making appropriate adjustments to the explicature. The ultimate result, an interpretation of the utterance, is a set of assumptions (taken to comprise the speaker's intended meaning) which meet the criterion of optimal relevance and are in an inferentially sound relationship with one another. The pragmatic process of *ad hoc* concept formation may result in a narrowing of denotation, e.g., the use of 'drink' to mean alcoholic drink, or a broadening, e.g., the use of 'flat' to describe a surface that is relatively free of bumps, or various combinations of narrowing and broadening, e.g., 'princess' (its encoded meaning entailing royal parentage) used to denote a haughty, pampered, demanding young woman, and so including some non-royal women and excluding some actual princesses (the well-behaved ones). Some of these new senses for a word become sufficiently frequently used and widespread as to become established senses of the word; they are stored in the lexicon with the word's other established senses and retrieved together with them when the word is accessed; in such instances, we have typical cases of 'semantic polysemy'. However, many such *ad hoc* concepts/senses are merely occasion-specific and transient.

As an example, consider the word 'mother', which can be used in the following three ways (among others), to refer to (a) X's biological mother, (b) X's adoptive mother (legal but not biological mother), and (c) the person with whom X feels a special bond of reciprocal affection (who may not be X's biological or adoptive mother, but someone who gave her the kind of nurturing that is normatively associated with a mother). Let us consider an RT account of how this third concept

1 I view Donnellan (1966) as the first theorist to extend the contribution of pragmatics (of the full-blown sort, i. e. geared to the recovery of a speaker's communicative intention) to the proposition expressed by a speaker, when he rejected both a semantic and an implicature account of the referential use of definite descriptions, suggesting instead that the attributive/referential use distinction is what he called a 'pragmatic ambiguity' making the proposition expressed either singular (referential) or general (attributive). Since that early prescient, albeit undeveloped, remark, many more instances of pragmatic ambiguity have been mooted (Carston, 2002; Recanati, 2004).

2 The distinction between decoding processes and inferential processes has been somewhat reconstrued (or at least relabelled) in recent years due, primarily, to revisions in the way in which 'inference' is understood, so that even linguistic decoding is construed as a kind of inferential process (Sperber and Wilson, 2015, and see Sperber, 2018 for informal discussion of a terminological shift from 'ostensive-inferential' communication to 'ostensive-interpretive' communication). However, the distinction between the two kinds of process involved in utterance interpretation remains untouched and, for the time being at least, I see no harm in talking of 'the code', and of 'encoding' and 'decoding', taking these terms to concern the role(s) of the language system in linguistic communication, although I will suggest below that this code is a kind of hybrid, comprising two quite architecturally distinct parts.

expressed by 'mother' might be recovered in comprehending the following utterance:

1. I owe so much to my aunty Jane – she was my real mother

Assume the word 'mother' encodes (i.e., has as a conventionalized sense) the atomic concept MOTHER which provides a direct link to an 'encyclopedic entry' of assumptions/beliefs about mothers, including the following (and much more):

a. A mother is a female parent [with further information about biological mothers, adoptive mothers, step-mothers, surrogate mothers, etc.].
b. A mother is expected to provide the love and nurturing that ensures the child thrives physically and psychologically.
c. A mother may be controlling and manipulative in ways that are not beneficial to a child.

Some elements of encyclopedic information are more accessible (more highly activated) than others, depending on the content of the rest of the utterance and the specifics of the occasion of use. For the current example, the most highly activated items of information are likely to be those in (b), which are then used as contextual assumptions/premises in deriving *cognitive implications* (e.g., Jane gave the speaker the love and nurturing expected of a good mother; this was highly beneficial to her physical and emotional development, etc.), which, in turn, via a mechanism of 'mutual parallel adjustment' of explicit content, contextual assumptions and cognitive implications, modulates the concept expressed/communicated by the word, yielding an *ad hoc* concept MOTHER*, a concept whose denotation is both broader than the encoded concept MOTHER, as it includes people who are not the female parent of a child but have given the child a kind of motherly nurturing, and also narrower in that it excludes negligent mothers (who are female parents). This inferential process stops when context-specific expectations of relevance (formed on the basis of the presumption of 'optimal relevance' conveyed by all utterances) are satisfied. The *ad hoc* concept/sense that is inferred is a constituent of the explicature of the utterance, taking the place of the decoded concept MOTHER.[3]

The utility of this account in explaining cases of word meaning variation in other fields has been demonstrated recently by several applications in the philosophy of language. I focus here on one of these, as developed by Baumgartner (2022, 2023), who discusses so-called 'dual character concepts' (DCCs), that is, concepts that have both a descriptive dimension and a normative dimension. Standard cases discussed in the philosophical literature are 'poet', 'artist', 'philosopher', 'scientist', 'friend', 'soldier', 'woman', 'man'. An attested case of the last of these is the statement 'Hillary Clinton is the only man in the Obama administration', where 'man' is clearly not being used descriptively (to mean 'male, human, adult') but normatively, that is, to pick out properties which, according to a (now largely discredited) social stereotype, are expected of a 'real man': psychological strength and courage, forcefulness, steadiness in the face of adversity, etc. (see Leslie, 2015 for extensive discussion of this example). The descriptive concept MAN and the normative concept MAN* expressed here are what the philosophers term 'fully dissociative' in that they set up two distinct categories: someone may be a man descriptively but not normatively (i.e., an adult male who lacks the normative properties of mental strength, courage, etc.), and someone else may be a man only normatively (as Hillary Clinton is claimed to be in the utterance above). An RT account of the latter concept MAN* would be essentially the same as that given above for MOTHER*, using the social stereotype of a 'real man' to derive cognitive implications about the person so described, from which, in turn, the *ad hoc* concept is derived, a concept whose denotation is both narrower in some respects and broader in others than the descriptive concept (hence the noted dissociation between the categories they denote).

Most of those who have analyzed the DCC phenomenon take a semantic view, maintaining that those words which have this dual character (e.g., 'philosopher', 'artist', 'scientist', 'mother', 'man', 'woman') are cases of lexical polysemy, both senses being established across a population of users and stored in their mental lexicon (e.g., Leslie, 2015, p. 120). However, as Baumgartner notes, the virtue of the pragmatic account as given above is that it can explain a much wider range of cases than the lexical semantic view, which is restricted to those that have become conventionalized. It is certainly possible that some are now cases of semantic polysemy, e.g., 'man' and 'mother' in the normative senses discussed above, perhaps also 'philosopher' in the sense of a person who typically seeks answers to difficult questions about meaning or ethical issues via rational thought/argumentation, whether or not that person is a professional philosopher (Baumgartner, 2022). However, the pragmatic inferential account gives us both an explanation of how these established normative senses arose in the first place, and an account of cases that are not lexicalized and/or trade on normative values that are not public or established, but are themselves *ad hoc* and contextual. For a possible case of the latter, imagine the following: there is a family, the Hansens, the mother of whom emphasizes to her children that they should stay positive, calm and cheerful, even when difficult or upsetting things happen to them; while most of the family manage to comply with this 'norm' most of the time, the youngest child, Billy, tends to be moody and morose; next door lives his best friend, Joey Wilson, who is a happy-go-lucky boy. One day, Mrs. Hansen admonishes Billy, saying: 'Joey is more of a Hansen than you are', meaning, of course, that Joey, who is not a Hansen but a Wilson, has the (normative) characteristics of a Hansen family member: he is a HANSEN*. As Baumgartner says, the lexical pragmatic approach (in terms of *ad hoc* concepts) can account for the full range of cases (whether *ad hoc* and transient, established and lexicalized, or somewhere in between), while the lexical polysemy account applies only to the lexicalized (conventionalized) cases.

My aims in this section have been: (a) to briefly describe the relevance-based account of *ad hoc* word meaning creation as a pragmatic contribution to utterance comprehension and a significant source of polysemy, and (b) to show, via exposition of Baumgartner's work on dual character concepts, the potential utility of this account in helping to explain certain cases of multiple word meaning which

3 There are many more detailed discussions and exemplifications of *ad hoc* concept construction in the relevance-theoretic literature (Sperber and Wilson, 1998, 2008; Carston, 2002, 2019, 2021; Wilson and Sperber, 2002, 2004; Wilson and Carston, 2007; Falkum, 2017).

are central to debates in the philosophy of language.[4] In the next subsection, I move to a different kind of lexical pragmatic creativity, that is, the coining of new (*ad hoc*) words in the process of online communication/interpretation.

## 2.2. Lexical innovation: *ad hoc* words and the role of metonymic associations

New words coined on the fly in communication (as opposed to via offline stipulation) may take various forms, including (a) cases of standard word formation involving regular processes of affixation as in 'detector-ist', 'expir-ation', 'burglar-ize', 'worst-est'; (b) blends, which take parts of two distinct words and form a composite, e.g., 'brunch', 'motel', 'franglais', 'blizzaster'; (c) conversions, e.g., the verbs 'to favorite', 'to laser', 'to lawn', 'to prodigy' based on pre-existing phonologically identical nouns. There is a syntactic-semantic story to be told about how these new coinages acquire their compositional meanings and a relevance-theoretic pragmatic story to be told about how, for those that have them, they acquire non-compositional meanings. The syntax and pragmatics of the affixation cases are addressed in Section 3, but here I focus on 'conversions' (and specifically denominal verbs), which are distinctive in that there is no phonological difference between them and the nouns on which they are based.

Here is a sample of the phenomenon at issue: (2a)-(2b) are attested new(ish) *ad hoc* cases, (2c)-(2e) are more familiar, with two clearly distinct uses of the verbs in the (d/d') and (e/e') cases, and those in (2f) are fully established/conventionalized:

2. a.'I'm trying to *room* all the talks in the same building' (conference organizer).
b. 'Vasko Vassilev *prodigied* his way through the *Carmen Fantasia*' (Alan Rickman).
c. 'The prisoner *houdinied* out of the top security jail'.
d. 'The factory *sirened* midday'.
d'. 'The police *sirened* the Porsche to a stop'.
e. 'The boy *porched* the newspaper' [threw X onto a porch].
e'. 'The developer *porched* all the bungalows' [added a porch onto X].
f. to hammer (a nail, a box flat), to shell (walnuts), to starch (shirts), to dust (the corners of the room; the cake with cinnamon); to treasure (our time together); to bike, to bus, to jet; …

On a traditional linguistic view, conversions are cases of derivational morphology, essentially the same as the move from the noun 'standard' to the verb 'standard-ize', or from 'code' to 'cod-ify', but with a zero (phonologically empty) affix. However, advocates of this view have often noted, with some unease, the extensive range of meanings that conversion verbs can have, meanings that are unsystematic and unpredictable, unlike that of typical cases of affixation. Consider, for instance, the very different kinds of interpretation (and relation between verb and parent noun) of 'to room', 'to prodigy', 'to porch', 'to siren', and 'to dust'. In their ground-breaking study of nouns 'surfacing as verbs', Clark and Clark (1979) treated them as cases of *lexical innovation*, new words coined on the fly in communication, whose meanings are highly context-sensitive, with only the very general linguistic constraint that they are verbs.

So these spontaneously coined denominal verbs require a pragmatic explanation, in which the encyclopedic information which comes with the parent noun, e.g., about porches in the cases of (e) and (e'), plays a key role, along with readily accessible contextual information, e.g., about boys delivering newspapers or developers building houses for (e) and (e'). In his study of conversions (both noun to verb and verb to noun), Bauer (2018) takes this pragmatic account one step further, maintaining that they are *metonymic shifts* made by speakers in communication. As he puts it, they are typical of figurative interpretations in being 'unpredictable and unrestricted' (Bauer, 2018, p. 180) and the relations between the meanings of the parent noun/verb and the derived verb/noun are typical of metonymic associations, e.g., *location for action/event* as in '*porch* the newspapers', *attribute for behavior* as in '*prodigy* the *Carmen Fantasia*', *person for behavior* as in '*houdini* out of the cell', *instrument for action* as in '*siren* the car to a stop'.

However, there is a notable departure here from standard cases of metonymy, as reflected in the following definition of metonymy: 'a figure of speech involving substitution of the name of an associated attribute or adjunct for that of the thing meant' (OED). That is, metonyms typically involve the use of a noun to refer to an associated entity, person or thing rather than to an action or process, and so *do not involve a change of syntactic category*. For instance, 'a farm hand', 'the city suits', 'the crown', 'Downing Street', 'the ham sandwich', and a wide range of semi-regular cases: e.g. *container for contents* (e.g., 'He drank the whole bottle'); *creator for work* (e.g., 'I've read Dickens'); *place for event* (e.g., 'Waterloo', 'Vietnam', 'Woodstock'); *animal for meat*, etc. The relation is often described as one of 'contiguity' (spatial, temporal or casual/resultative) between *things* in the world (distinguishing it from other cases of non-literal use: resemblance for metaphor and antonymy for irony).

Conversions do not seem to fit this standard definition of metonymy, and more generally, figures of speech (e.g., hyperbole, metaphor, irony) do not usually involve a change of word category (creating a verb from a noun, or vice versa). It might seem then that in these standard cases of nominal metonymy, e.g., 'hand', 'suit', Downing Street', 'ham sandwich', what we have is just another instance of *ad hoc* concept construction, as discussed in the previous section, where the word or phrase is given a new meaning (which may become established over time giving rise to semantic polysemy). However, in recent work within relevance theory on these standard cases of nominal metonymy, Wilson and Falkum (2015, 2020) have argued that, in fact, 'metonyms arise as motivated neologisms' i.e. metonymic uses are spontaneous pragmatic processes of *new word coinage*. On their pragmatic account, the new word (specifically its meaning, as its phonology is a given) is inferred from accessible information in the encyclopedic entry of the input noun (e.g., 'suit', 'hand', 'ham sandwich') and information in the wider discourse context, guided by the prevailing relevance-theoretic comprehension procedure. Importantly, they maintain that this is different from meaning

4 For another, quite different, application of *ad hoc* concepts within the philosophy of language, see Liu (2023), who argues that phenomenal 'what-it's-like' concepts typically originate as *ad hoc* concepts (pragmatically narrowed) and that this presents a challenge for work in experimental philosophy that tests whether laypeople grasp these concepts and draws conclusions from their apparent failure to do so.

modulation (narrowing/broadening/metaphor) because the encoded concept/sense (e.g., HAND, SUIT, HAM SANDWICH) and the new *ad hoc* (metonymic) concept/sense (e.g., HAND*, SUIT*, HAM SANDWICH*) do not share cognitive implications and their denotations are disjoint; the output is thus a different (*ad hoc*) word from the input word.[5]

What we see here is a nice convergence of independent work: Bauer's (2018) claim that conversions (new words made from existing words, e.g., denominal verbs) are instances of metonymy and Wilson and Falkum's (2015, 2020) position that standard nominal metonymies are motivated word coinages (denominal nouns). Putting these together, what we get is the view that when an existing word is used (without affixation or any other phonological change) to convey a metonymically associated meaning (or to refer to an associated entity/action/process in the world) a new word is thereby created, which may or may not involve a syntactic category change.[6,7] This applies equally to words that are more transparently complex because of their affixations, so, e.g., 'transmission' with its meaning of CAR'S GEARBOX looks like a case of *process for instrument* metonymy and 'reading' with its meaning of AN INTERPRETATION (as in 'His reading of the text was more allegorical than mine') is a *process for result* metonymy. If Wilson and Falkum are right, these are new words, distinct from the words 'transmission' meaning THE PROCESS OF TRANSMITTING and 'reading' meaning THE PROCESS OF READING, so they are further instances of denominal nouns. These morphologically complex cases are discussed in more detail in Section 3.1.

A worry about metonymy as a means of using existing words to generate new words/senses is that it seems to be very general and unconstrained, allowing users to take a word and form a new (phonologically identical) one whose meaning/content is in some sort of salient associative relation with the meaning/content of the existing one; as long as a speaker can be more or less sure that the association (spatial, temporal, resultative) is apparent to her audience, she is free to create the new word. However, metonymy just does seem to be an easy basic conceptual/pragmatic process. It arises spontaneously and cross-culturally very early on in children's communicative use (even pre-linguistically) and in their comprehension (well before they can comprehend metaphor) (Falkum, 2019; Köder and Falkum, 2020; Wilson and Falkum, 2020). Experiments testing people's appreciation of well-established polysemies find that they consider metonymically-related senses to be more closely related than cases involving 'resemblance' (narrowing, broadening, metaphor) (Klepousniotou and Baum, 2007). In fact, there is less 'semantic overlap' (i.e., sharing of features/properties) in metonymy than in the resemblance cases, related by pragmatic modulation, so the apprehended 'relatedness' must simply reflect the strength of the associative connection in people's minds. Klepousniotou et al. (2008) assume there must be some sort of 'core meaning' shared by established metonymies like the animal-meat (e.g., 'lamb') and institution-building (e.g., 'school', 'church') cases and that these are all 'literal' (rather than figurative) uses of the words involved (for discussion, see Carston, 2021). Whatever one may think of these assumptions, they indicate that metonymic associations (spatial/temporal/resultative contiguities in the world as apprehended by us) are quite psychologically basic. Although metonymy is usually seen as less interesting (and certainly less beautiful) than metaphor, it may be that, in certain ways, it is more fundamental to our cognitive and communicative lives.[8]

Note that a significant consequence of the view that metonymic conversion is a means of creating new words is that polysemy must cross syntactic categories, that is, a family of closely related senses may be spread across nouns and verbs, e.g., 'porch', 'starch', 'houdini', 'jet', 'prodigy', and even across nouns/verbs/adjectives, e.g., 'stone', 'back' (Carston, 2019). What these words share is their phonology and, more crucially (since homonyms also share phonology), a *root*, which can be notated as follows: √stone, √porch, √houdini, √prodigy, etc. So, it is really roots rather than words that track polysemy (i.e., families of interrelated senses). Yet words seem to be highly salient to ordinary language users (Julien, 2007), and it is words, rather than roots, that are employed as our minimal communication units (one-word utterances) and are logged in our pragmatic lexicons (see Section 3.3). Some current syntacticians maintain that *words have no status in the grammar*, so, e.g., 'nationalize', 'solidarity' are phrasal structures, as are 'siren', 'porch', and even apparently simple words like 'cat' and 'run'

5 I should point out that the ideas of Wilson and Falkum that I draw on here come from a series of conference presentations, and may not represent their final position when the account is published. In her work on referential metonymy, Bowerman (2019, 2021) takes a somewhat different but possibly compatible position, according to which a metonymic use of a word or a phrase (e.g., 'the ham sandwich') is a *repurposing* of the literal meaning of the expression in order to facilitate the interpreter's access to a novel referent (e.g., a customer who ordered a ham sandwich), on the basis of a contextually salient relation between its literal referent (e.g., an actual ham sandwich) and the speaker's target.

6 As noted by Bauer (2018), others have also suggested that conversions are cases of metonymy or discussed the pros and cons of the idea (e.g., Dirven, 1999; Cetnarowska, n.d.).

7 One of the reviewers raised the interesting question whether new words (derived from phonologically identical ones) must always involve a metonymic relation between senses and suggested that the relation might sometimes be metaphorical. They further suggested that the verb 'houdini' might be such a case, being based on a resemblance between our concept of the man Houdini and our concept of the action of escaping in an incredible way. While I am doubtful about this specific example (it does not seem that the individual concept encoded by the name 'Houdini' is modulated into a verbal (action) concept meaning 'to escape incredibly'), I would agree that the senses of some new words may bear a metaphorical relation to their origin word's sense – this is something that definitely needs further consideration. For now, I follow Wilson & Falkum's (unpublished) work on this and take the following conditional position: if the pragmatic relation between a new/*ad hoc* sense and an existing sense of a single phonological form is metonymic, then the new sense is the meaning of a new word (or new communication unit).

8 It has been suggested by Dan Sperber (2017) that non-human primates may also use metonymy. Here he draws on experimental work with apes by Bohn et al. (2015), in which chimps, bonobos and baboons pointed through a wire mesh to an empty plate in order to request a certain preferred kind of food (grapes) which they had previously received on that plate, which looks very much like a case of *container for contents* metonymy. This may be an evolutionary precursor to the human use of metonymy, but it is rather different from human metonymy because only humans have words, that is, phonological units with syntactic structure which couple up with (families of) discrete atomic concepts/senses (Pettito, 2005).

(according to Marantz (1997), Borer (2015), Harley (2014), and many others), for which the basic elements are roots (and functors, including categorizers). As Acquaviva (2022, p. 283) says: root-based syntax 'may model the distinction between polysemy and homonymy in formal terms, so only polysemous words share the very same syntactic root', so, e.g., there is a single root for the noun/verb/adjective 'stone', with their pragmatically interrelated meanings, but two roots for homonyms like 'bank' and 'bug', which have two unrelated families of meanings. The syntactic side of conversions and of other new words, and the importance of roots, are discussed further in the next section.

## 3. The language faculty and ostensive communication

### 3.1. 'Words': syntax and pragmatics

According to the root-based approach to syntax, touched on above in the discussion of conversions and cross-categorial polysemy, words have no formal or theoretical status in the grammar; they are phrasal structures and so, like all phrasal structures, they have a compositional semantics, which is a function of the meanings of the basic parts and their structural combination. On some views (e.g., Panagiotidis, 2014; Borer, 2017), a root is nothing more than an index or address tracking its occurrence across categories, so it is meaningless as well as categoryless; only once it has been conflated with a syntactic categorizer is it assigned a meaning, e.g., [$_N$ √form], [$_V$ √houdini], [$_{Adj}$ √stone]; this is the first level of content, on the basis of which the compositional meaning of a more complex word (e.g., 'formation', 'stonily', 'adorable', 'houdinify') is generated.

However, as widely noted, there is a glaring issue for this single syntactic engine approach to word structure: many of those phrasal entities that we think of as words have a *non-compositional (idiosyncratic, unpredictable) meaning*. Examples abound: 'reactionary' meaning BACKWARD-LOOKING, 'transmission' meaning GEARBOX OF A CAR, 'flakey' meaning UNRELIABLE, 'execution' meaning STATE-SANCTIONED KILLING, 'demonstration' meaning ORGANIZED MASS PROTEST, 'naturalize' meaning MAKE SOMEONE A CITIZEN OF A COUNTRY, 'liquidate' meaning KILL SOMEONE (VIOLENTLY), 'recital' meaning SOLO CONCERT … There are two points to note here: (a) Each of these words also (inevitably) has a compositional meaning, although there is considerable variation in the current usage of these meanings (e.g., while the compositional meaning 'TRANSMIT + tion' is widely used, the compositional meanings 'RECITE + al' and 'REACTION + ary' are much less so); (b) The non-compositional (idiosyncratic) meanings are not completely unrelated to the corresponding compositional meanings (or to the meanings of other words with same root). What lies behind this relatedness is, as already noted in Section 2, the fact that the very same pragmatic processes of meaning/sense modulation and metonymic word coinage, typically discussed within relevance theory only with regard to monomorphemic words, apply equally to these more structurally complex cases, a point exemplified further below.

The key issue here for the syntactic treatment of word structure is that it does not account for the non-compositional meanings that complex words can have. Of course, the advocates of this approach to word structure are well aware of the issue and some have developed explanations for *why* and *when* non-compositional meaning is possible, although not for the particular meanings that arise (which, I maintain, is the job of pragmatics). Their general idea is that there are specific 'syntactic domains' within which non-compositional (atomic, idiosyncratic, special, unpredictable) meaning/content can emerge. So 'recital', 'naturalize', 'reactionary' have a specific *kind* of syntactic structure which, although it has a compositional meaning (like all syntactic structures), allows for (but does not require) assignment of a special (non-compositional) meaning. I cannot begin to assess here the relative strengths and weaknesses of the various proposals, which are often highly technical and developed within different syntactic frameworks. What follows is a very simplified indication of one of the best-developed accounts, that of Borer (2013a,b, 2014).

Borer's syntax is a 'constructionist' theory: the grammar generates syntactic templates (event structures) into which roots are inserted, e.g., √dog or √stone is inserted into a structure within which it becomes noun-equivalent, verb-equivalent, or adjective-equivalent, depending on its position within the structure. As well as roots, there is another kind of basic element in the system, namely 'functors' (which include tense, aspect and number indicators, determiners, and categorizing affixes such as '-ize', '-ary', '-tion', etc). Borer distinguishes two different kinds of functors and these play crucially different roles in her account of syntactic domains of non-compositional meaning (or Content, as she calls it). These are (1) *C-functors*, i.e., categorizers, e.g., *n*ominal, *v*erbal, *a*djectival, which may (but need not) be realized phonologically by various affixes, and (2) *S-functors*, which project further levels of structure; these include the determiners (e.g., 'the', 'those'), count/mass and number (singular/plural) indicators for nominal structures, and tense and aspect for verbal structures.

Categorizers (C-functors) allow non-compositional meaning assignment at multiple levels, so, for instance, in the structure 'the [{([√nature $_N$] al $_A$) ize $_V$} ation $_N$]', Content can be assigned at each of the structural domains headed by N, V, or A, and, as noted above, the domain delimited by V here has, in fact, received a non-compositional (idiosyncratic) meaning/content: MAKE SOMEONE A CITIZEN OF A COUNTRY. Structures headed by S-functors do not allow this:

3. a. Tense phrases: 'jump-ed' – meaning must be compositional.
   b. Number phrases: 'book-s' – meaning must be compositional.
   c. Determiner phrases: 'the/that/my book' – meaning must be compositional.
   d. AS-nominals (which inherit the *A*rgument *S*tructure of the verb from which they are derived): e.g. 'destruction' (of the city by the barbarians in a single day), 'teaching' (of the physics class by Mary).

C-functors indicate structure points at which Content (that is, non-compositional meaning) can be assigned. On Borer's (2013a, 2014) account, these are points at which there is a search of what she calls the 'Encyclopedia' for a matching content. The 'Encyclopedia' is not a component of the grammar, but rather lies within the Chomskyan conceptual-intentional (semantic-pragmatic) systems with which the syntactic engine interfaces, and it is the locus of stored non-compositional meanings. It is akin to (though by no means identical to) my conception of the communicative lexicon, discussed below in Section 3.3.

A striking piece of evidence in support of Borer's account comes from the behavior of two different kinds of verb-derived

nominals, examples of which have already been briefly mentioned: those that can take a non-compositional meaning, e.g., 'transmission', 'recital', 'referral', 'revolution', 'solicitor' (known as R-nominals), and those that cannot, as in 3(d), (known as A-S nominals). To make this clearer, here are instances of the two kinds of case, where each member of the pair, (a) and (b/c), has been derived from the same verb and is phonologically identical:

4. a. The transmission of news from Ukraine by the BBC (for several hours…).
[*cf.* The BBC transmitted news from Ukraine for several hours.]
b. The car's transmission [= gearbox] is in good condition.
c. * The car's transmission by Nissan for several years …
[cannot mean: *the transmission (= gearbox) as made* by Nissan for several years].

5. a. The referral of Mary by her doctor to a rheumatologist.
[*cf.* The doctor referred Mary to a rheumatologist].
b. The referral [= person referred] left the consultant's room feeling reassured.
c. * The referral by her doctor to a rheumatologist [has arrived for her appointment].

In each of these cases, the (a) version has inherited the argument structure of the verb from which it is derived and its meaning is compositional, while the (b) version has a non-compositional meaning and it does not allow the verbal arguments, as shown in (c), in each case. That is, we find a correlation of properties here: *R-nominals* can have special (non-compositional meaning) and do not take argument structure; *AS-nominals* have argument structure (whether explicit or left implicit) and cannot have special (non-compositional meaning). Members of each pair are derived from the same verb, [transmit], [refer], and are phonologically identical, so are distinguished by syntax alone. Here are their syntactic structures (simplified):

6. R-nominals (can have non-comp meaning): ($_N$ -tion [$_V$ transmit])
AS-nominals (cannot have non-comp meaning):
[$_N$ -tion ([$_{F2}$ subj [$_{F1}$ obj [$_V$ transmit]]])].

The R-nominal is headed by a categorizer with no intervening S-functors, so marks out a structural domain which allows assignment of Content. The AS-nominal, on the other hand, contains what is abbreviated here as functional structures F1 and F2 (the subject/object arguments, e.g., 'the BBC/the news', 'the doctor/to a rheumatologist'), so is replete with S-functors, which block assignment of Content (non-compositional meaning).

Assuming, then, that Borer's account is well-grounded, a hypothesis, whose confirmation would be very pleasing for the picture I am drawing here, is that the syntactic structures defined by these domains are typically what the language user perceives as 'words' and which can, therefore, be the basis of the kind of pragmatic lexical modulation processes that were discussed in Section 2. This seems plausible but needs empirical support. If it proves to be right, these domains provide the necessary link between the formal computational system and what I call the pragmatic or communicational (user-based) lexicon.

To end this section, let me indicate, with some more examples, the ways in which the non-compositional meanings of some of the morphologically complex words discussed above mesh with the pragmatic account of meaning modulation (narrowing/broadening) and new word coinage. I leave the specifics of plausible contexts for these meaning creations to the imagination of the reader. The verb 'naturalize' with the non-compositional meaning NATURALIZE (= make a foreigner into citizen of a country) is a pragmatic narrowing of the more general compositional meaning [NATURAL + ize] roughly paraphrasable as 'to make natural'. Such narrowings are common in specific contexts in which jargon terms arise: e.g. 'transformation', used for a kind of grammatical operation in linguistics, and 'transference' used in psychoanalysis for a particular psychological process, both narrowings of the general compositional meaning of the structures involved. These are both R-nominals (and have AS-nominal counterparts). Something a bit different is going on with the following cases suffixed by '-ing': 'reading' with the non-compositional meaning READING (= an interpretation), as in 'His reading of the novel was highly allegorical', and 'teaching' with the non-compositional meaning TEACHING (= a set of ideas/a lesson), as in 'She was profoundly influenced by Buddhist teachings'. These seem best analyzed as involving a *metonymic shift*, given that arriving at an interpretation of a text is typically a *result* of a process of reading (= READ + ing) that text and a set of ideas is typically a *result* of someone's teaching (= TEACH + ing) them, a standard metonymic relation, according to Bauer (2018), and therefore, if the account of metonymy given in Section 2 (Wilson and Falkum, 2015, 2020) is right, these are new words, new communication units for users. In the case of 'reading', there appears to have been also a broadening of meaning in that one can have a reading not only of texts but also of situations and people: e.g. 'On my reading of the situation, we are doomed.' Furthermore, the verb 'read' itself seems to have acquired this meaning of 'interpret': 'As I read the situation, we are doomed', perhaps by some sort of back-formation process, and thus a compositional meaning of 'reading' (= interpretation) is reinstated. Finally, a similar sort of analysis of the non-compositional meaning of 'transmission' (= car's gearbox) can be given: a *narrowing* of denotation (to the specific kind of transmission that takes place in the engine of a car) and a metonymical transfer to the object responsible for this specific kind of transmission (the gearbox), creating a new word or communication unit for language users. For more detailed discussion, see Carston (2022).

Summing up: what I hope to have shown here is that a root-based syntactic account of word structure with C-functor defined domains of Content can be integrated with the relevance-based pragmatic account of how specific non-compositional meanings (atomic Contents) of words arise (by meaning modulation and metonymic transfer) to give a full and unified account of word meanings. The account applies equally to the overtly affixed cases discussed in this section, to so-called 'conversions'[9] such as the verbs 'porch', 'houdini',

9 I say 'so-called' conversions because on the constructionist account (Borer, 2013a, 2014), there is no 'conversion' process: e.g. the noun and verb pair 'hammer' simply arise from insertion of the root √hammer into two distinct formal structures, each of which is a domain for atomic meaning/Content. It is completely irrelevant to the syntactic system which of these, noun or verb, 'came first', in terms of its coinage and use by communicators and its storage in their lexicons.

'prodigy', and to seemingly simple words such as the noun 'cat', the verb 'put' and the adjective 'red'.[10]

## 3.2. Language faculty: narrow/broad; generative/stored; individual/social

According to Hauser et al. (2002), the broad folk notion of language is a mosaic of components, and fruitful investigation requires carving up this broad conception into tractable domains of study, separating out "questions concerning language as a communicative system and questions concerning the computations underlying this system, such as those underlying recursion." (Hauser et al., 2002, p. 1567). Generative linguists in the Chomskyan tradition focus on the latter, the narrow internal linguistic system (I-language), that is, syntax and its interfaces with conceptual-pragmatic capacities, on the one hand, and perceptual/articulatory systems, on the other. This computational system is not essentially an instrument of communication, although, as a matter of fact, it is widely and productively employed in communication, as enabled by its interfaces. Most psychologists, on the other hand, focus on language as a communication system, investigating the perceptual and cognitive processes that take place when we produce and comprehend linguistic utterances.

Words have a sort of double status, as they straddle the linguistic/communicative divide: they are phrasal entities generated by the syntax (language narrowly construed) – there are no words without syntax – but they are also salient as basic communicative units with shared meanings, some of which become conventionalized and so enter into the overlapping lexicons of particular groups of communicators, making them a socio-cultural component of language broadly construed. While formal linguists are interested in words as syntactic entities (with phonological and semantic properties), psychologists tend to focus on them as communicative units, investigating their activation, their retrieval/recognition, and their integration with other words in the course of utterance processing, often measuring the time course of their online production and comprehension.

These very different stances are more a matter of preferred focus than of incompatibility or rival positions – of course, we want an account of language as a faculty of the mind *and* an account of how it works when used in communication. However, the two different orientations can become incompatible (even antagonistic) when the discussion moves to the fundamental nature of language, what it is for, and its evolutionary origins. Some psychologists maintain that language just *is* a communication system and that it should be investigated as such, and be taken to have evolved (been selected) for that purpose. For instance, Vigliocco et al. (2014) decry: 'the (explicit or implicit) assumption that the object of investigation – language – can be properly and sufficiently addressed by ignoring other characteristics of face-to-face interactions: the communicative context in which language has evolved, in which it is learnt by children, and in which it is most often used', maintaining that language must be studied as a multimodal mix of speech, prosody, gesture and facial expression. This is, of course, directly at odds with the generative linguistic stance according to which the core property of human language is its recursive syntax, which is taken to be an entirely proper study in itself (Hauser et al., 2002; Berwick et al., 2013).

The 'purpose' of language and its evolutionary origins similarly divides these groups. Those with a communication-orientation maintain, albeit with important differences of detail among them, that it emerged in order to fulfil social-cooperative-communicative needs specific to humans (Tomasello, 2008; Vigliocco et al., 2014; Scott-Phillips, 2015). The syntax-oriented theorists point out that linguistic structures, with their gaps and long-distance dependencies, are not optimized for communication but rather for computational ease in thought (Chomsky, 2010), and that some easily interpretable structures are glaringly ungrammatical (e.g., 'Who did John call Mary and --'). On this 'biolinguistic' view, language (i.e., the capacity to recursively combine concepts) arose quite suddenly in the species via a rewiring of the brain, which conferred considerable fitness-enhancing advantage on the individual so endowed, enabling complex thought, understanding, and planning, a capacity then transmitted to offspring, and so coming to predominate. On this view, the use of language for communication is a subsequent and secondary development, a matter of linking the core linguistic capacity to sensorimotor systems required for its externalization (including the property of linearization, arguably not necessary for thought) for verbal utterance production and perception/comprehension.[11]

Where are relevance theorists situated in this debate? As pragmaticists, their focus is, of course, on communication, and on linguistic meaning as providing evidence of what the speaker intends to communicate, rather than on detailed investigation of the formal properties of language. However, there appears to be something of a divide between those relevance theorists who see language as a system whose raison d'être is communication, having evolved as an instrument of a pre-existing ostensive communicative capacity, which created a particularly favorable environment for the emergence of language (Sperber, 2000; Sperber and Origgi, 2010; Scott-Phillips, 2015), and those who find the Chomskyan story more promising, that is, that language (understood as the core recursive computational system) first effected a transformative change in our powers of thought, only secondarily being externalized and used in communication (Carston, 2015, 2023; Reboul, 2017). This a fascinating

10 Another very significant class of words, namely, compounds (e.g., 'skyscraper', 'flowerbed', 'eavesdrop', 'earmark') needs to be discussed in this regard. They are another case whose meaning seems clearly non-compositional (most strikingly so in the case of 'eavesdrop'), hence requiring a pragmatic account (see Bezuidenhout (2019) on the semantics/pragmatics of noun-noun compounds in English). In her study of noun-noun constructs in Hebrew, Borer (2013b) shows that, from the point of view of (non)-compositionality, there are two kinds of case such that, as with the nominals discussed above, 'it is the syntactic differences between them that give rise to distinct Content properties, with non-compositionality correlating with … the absence of functional structure' (Borer, 2013b, p. 205).

11 See Durrleman et al. (2022) for some intriguing empirical evidence from bilingual children with atypical development which they interpret as supporting the chomskyan position that: 'language, though useful for communication, is not sufficient for communication, and may arguably not have evolved primarily for communicative purposes [but for thought]' (Durrleman et al., 2022, p. 5).

and complex area, which I cannot pursue further here, but in this paper I am assuming the latter position. The next section addresses words as conventionalized communicative units, that is, as components of language broadly construed.

## 3.3. The communicational lexicon and polysemy

The way words work in our cognitive and communicative lives is very different from the way syntax works. First, as discussed in previous sections, word meanings are flexibly manipulated/adapted and new words are fashioned from existing words in order to express new concepts in ever-evolving contexts of communication. There is nothing comparable to this in the realm of syntax. Second, we continue to acquire new words throughout our lifetime, while our native language syntax is essentially in place and fixed by the age of five or six. Admittedly, young children learn words at a remarkably fast rate (several per day at the peak of acquisition), but they are also learning vast numbers of new facts at the same age. In line with this, the evidence indicates that this acquisition process is not achieved via a dedicated cognitive system as is the case for syntax, but rather via more general cognitive processes of learning and memory which are also employed in the acquisition of new facts.[12] For instance, Markson and Bloom (1997) report studies in which children aged 3.7 and 4.3 were (like adults) as good at learning and remembering an arbitrary linguistically presented fact about a new unfamiliar object (e.g., 'My uncle gave it to me') as they were at remembering its name (e.g., a new word 'koba'), and this was so even when the new arbitrary fact contained a novel word (e.g., 'It came from a place called Koba').[13]

As noted in the previous section, although words are syntactic entities, they have no privileged status in the narrow language faculty (syntax and its interfaces); they are simply one of the many phrasal structures generated by the syntax, whose basic units are roots and functors. For the ordinary language user, however, words are highly salient as basic units of communication[14] and are stored in a lexicon, from which they are retrieved (with their families of related senses) in linguistic communication and comprehension, along with other linguistic phrases that have become well-established (conventionalized) and are accessed as a whole: e.g. idioms such as 'spill the beans', 'trip the light fantastic' and frozen forms such as 'in cahoots with', 'kith and kin'. An individual's communicational lexicon is a result of her communication history and consists of phonologically spelt-out forms and conceptual meanings, which can accrue many cultural/personal associations. This lexicon is a *performance system*, in Chomsky's terms; it is a component of the language faculty only on a broad construal, lying outside the narrow linguistic 'competence' system and arising from socio-cultural processes of communication and conventionalization; it registers properties of words (and their senses) like frequency of use, which are irrelevant to the syntactic system, but are reflected in the processes of word recognition, retrieval, priming and comprehension measured in online psycholinguistic experiments.[15]

If the relevance-based pragmatic story told in the earlier sections is right, new *ad hoc* word senses/meanings are being fashioned in context all the time (via the concept modulation process that results in more specific and/or broader senses) and new *ad hoc* words are being coined (via metonymic associative processes), but only some relatively small subset of these becomes sufficiently well-established so as to enter a user's mental lexicon. We need an account of the socio-cultural process(es) that result in the conventionalization of words and senses (setting aside those that come into being via authoritative stipulation); words as communication units (rather than syntactic entities) are, arguably, cultural phenomena and so are to be explained in similar terms as the evolution of other stable cultural items. The challenge here is to explain how the language use of individuals leads to group or population-wide communicative conventions. This is not an issue I can pursue here, except to mention briefly the interesting 'epidemiological' framework initiated by Dan Sperber. According to this approach, cultural phenomena, including words, are to be explained as the cumulative effect of multiple processes taking place within and between individual members of a population, that is, causal chains of mental and public representations, imperfectly copied from one token to the next but sufficiently similar to constitute a recognizable and stable type (Claidière et al., 2014).[16]

There are many questions about the nature of the entries for words in this communicational lexicon; focusing on the meaning side, I assume a word (that is, a phonological form classified as a noun,

12 As Markson and Bloom (1997, p. 815) note: 'Children are much better at learning phonology, morphology and syntax than adults, consistent with the notion of a biological specialization for these aspects of language'. That is, there appears to be a critical period for acquisition of the structural components of language, but not for the acquisition of words.

13 This is not to say, however, that children do not appreciate a difference between words for things and facts about things ('things' used broadly here to cover not just entities, but also processes, activities and events in the world); they certainly do, especially concerning the conventionality of word meanings, and key differences between kinds of facts: generalizations ('Dogs bark') and one-off facts ('Uncle John gave me this'), only the former playing a useful role in categorization (as do words). See Tippenhauer and Saylor (2019) for a balanced overview of the ways in which words and facts, and their learning, overlap and differ.

14 Julien (2007), who argues strongly against words having any scientific or theoretical reality, nevertheless recognizes their psychological reality to us as language users, which she says is probably due to their distributional properties: 'since words are the minimal morpheme strings that can be used as utterances and that may be permuted more or less freely, words are the minimal linguistic units that speakers can manipulate consciously. It is therefore no surprise that speakers are generally aware of words' (Julien, 2007, p. 83).

15 Another important property of our lexicons is discussed by Clark (1998), which is that it includes various 'communal sublexicons', that is, we index or tag conventionalized senses of words to the specific community for whom we take them to be conventional (economists, linguists, Londoners, football aficionados, etc.) and deploy them accordingly.

16 Another 'pragmatics first' approach is that of Christiansen and Chater (2022), who show how the gradual conventionalization of gestures/signals in the game of charades can capture, in miniature, some crucial aspects of the cultural evolution of language, specifically words. As with the epidemiological approach, this seems promising as an account of the establishment of words and senses (basic units of human communication), but less so for the much more rigid structures of syntax that speakers seldom innovate with.

verb, adjective, etc. but with no more syntactic information than that) includes its family of related (established) senses. Most words are polysemous: consider, for instance, the verb 'run' in 'run a mile', 'run a business', 'run a meeting', 'run for president', the noun 'line' in 'a line on a page', 'a line of a face', 'a line of washing', 'a line of work' and the adjective 'shallow' in 'shallow water', 'shallow valley', 'shallow bore', 'shallow thought', etc. As noted by Wittgenstein long ago, in a discussion of the word 'game', there is no common definitional core (contrary to some current claims in the psychology literature) to these related senses; they are the result of chains of pragmatic inference, which take senses off in various directions dependent on the contexts in which they arise. It may be that the best way to think of how these sense families are represented is as a network of connected nodes, with proximate nodes representing closely related senses, and nodes separated by multiple nodes much more distantly related (Langacker, 1991; Recanati, 2017). Note that homonyms (e.g., 'bank', 'bug') will comprise two distinct entries in the lexicon each with its own network of related senses.[17]

The same account applies to noun-verb 'conversions', with their metonymically-derived non-compositional meanings, e.g., 'dust' (remove dust), 'dust' (sprinkle (e.g., sugar on a cake)), 'porch' (throw something on a porch), 'porch' (add a porch to a building), 'houdini' (make an incredible escape), and to other morphologically complex words, e.g., 'reactionary', 'naturalize', 'recital', 'transmission', 'detectorist', 'flakey'. They appear in the communicational lexicon with their non-compositional (atomic, idiosyncratic) senses. As noted in Section 3.1, they each also have a compositional meaning, composed from the sense assigned to their smallest domain of content, e.g., [$\sqrt{}$nature + n], [$\sqrt{}$recite + v], [$\sqrt{}$detect + v], plus the further levels of categorization in their structure. This compositional meaning can be seen as a function of syntax and the relatively rigid semantics of the affixes ('-ary', '-ize', '-ist', etc.), so it is predictable, not idiosyncratic, and on those grounds need not be listed in the lexicon. However, if that meaning has become conventionalized (do compositional meanings ever become conventionalized?), it might be listed along with the non-compositional meanings of the word; this remains an open question, not to be decided simply on grounds of theoretical economy (Carston, 2021, 2022). Either way, we have here further cases of polysemy, whose source is a combination of syntax and pragmatics.

What then is the language code, what is it that is decoded (albeit by a process of 'linguistic inference', according to Fodor, 1983 and Sperber, 2018; see footnote 2)? Its outputs seem to be mental representations which are a product of both the narrow language faculty (syntax) and components of its interfaces, of which what I call the communicational lexicon is most relevant here. Creative use and pragmatics (= interpretive inference) are what we do with these decoded outputs. That is, from the point of view of utterance comprehension, the code consists of two parts: syntax (the narrow 'linguistic' faculty) and lexicon (a part of our conceptual-intentional or semantic/pragmatic systems).

As already noted, the communicational lexicon is a component of the language faculty only as broadly conceived; it lies outside the formal computational linguistic system and is a store of communication units with non-compositional meanings. The form-meaning (syntax-pragmatics) divide emphasized in the previous sections, is real and can lead to dissociations in children's development, as shown by quite a wealth of empirical evidence, some of which is briefly surveyed in the next section.

# 4. The language-communication divide: empirical evidence

## 4.1. Autism and the form/meaning (syntax/pragmatics) divide

It is widely agreed and robustly attested that autistic children[18] are impaired in their social interactions and in communication/pragmatics (see, e.g., Tager-Flusberg et al., 2005; Kissine, 2021), while the state of their specifically linguistic abilities is much less clear. Given the pragmatic account of (non-compositional) word meanings presented in this paper and in more detail in Carston (2022), and the more general thesis of a divide between the formal (morpho-syntactic) and the conceptual (word meaning) components of language, several predictions or hypotheses about autistic people's abilities with words seem to arise: (a) inferring new, *ad hoc* (non-compositional) meanings in context is likely to be difficult, especially if the word already has an established meaning in the autistic person's lexicon, and so (b) their lexicons will contain little polysemy (i.e., networks of established related senses), while (c) the formal syntactic aspects of words may pose little difficulty, that is, acquisition of complex structures like 'formation', 'amplifier', 'demonstrate', 'transmission'.

Although it must be acknowledged that the evidence is patchy and any conclusions are tentative at best, I will survey some studies of autistic children that I believe point in the direction of a dissociation between their acquisition of formal aspects of language, on the one hand, and their grasp of conceptual word meanings and building of a communicative lexicon, on the other. As outlined in earlier sections, my general thesis is that the meaning components of our mental lexicons are fundamentally a product of communication, with many now established meanings of words having originated from processes of online pragmatic inference which take as input an existing word meaning and derive a new (*ad hoc*) contextually relevant meaning; this is a major source of the widespread polysemy of the communication units we deploy. If this is right, we might reasonably expect autistic children (and youth), who are known to have difficulty with flexible pragmatic word use, to exhibit concomitant difficulties in building a

17 Based on work in psycholinguistics on production (e.g., picture-naming tasks), Ramchand (2022) suggests that lexical entries are 'hubs' or 'lemmas', in the sense of Levelt (1999), that is, they house all inflectional forms of the 'same' lexeme (e.g., singular and plural for nouns; tense/aspect for verbs), while distinct derivational items based on a single stem are distinct words with distinct lexical hubs (e.g., 'form', 'formation', 'formative').

18 Alerted to current guidelines by one of the reviewers, I have changed my earlier use of the expression 'children with autistic spectrum disorder (ASD)' to 'autistic children' (except when quoting others). However, it is clear from the literature on autism that the people to whom the adjective 'autistic' is applied vary considerably in the degree to which they manifest associated abilities and disabilities, so the word 'spectrum' seems appropriate and helpful.

communicative lexicon, at least one that resembles that of typically developing (TD) children.

In a recent experimental study, Floyd et al. (2021) report that 'children on the autism spectrum are challenged by complex word meanings' (p. 2543); more specifically, they showed that autistic youngsters (aged 7–14) were shown not to have the facility with polysemy that TD children have. While the latter find it significantly easier to learn multiple related meanings of a word (polysemy) than to learn multiple unrelated meanings for a single word form (homonymy), the autistic group showed no difference between the two conditions. Floyd et al. conclude that polysemous words present a challenge to autistic children, and that they may benefit from interventions designed to help them 'to recognize that a word witnessed in a particular context with a particular meaning can also be used in a different context with a related but distinct meaning.' (Floyd et al., 2021, p. 2547).[19] These results and the conclusions drawn from them mesh well with the account of the communicative lexicon that I have given, from which a prediction of difficulty with building polysemy into word meanings follows from a more general difficulty in allowing for the kind of flexibility of word meaning required to form contextually relevant on-the-fly *ad hoc* word meanings.[20] However, there was, of course, no comparison being made here with these children's formal linguistic abilities, so we need to look elsewhere to see if there is support for the position that the development of grammar may follow a different trajectory, coming from what is, in effect, a different source (the narrow computational linguistic capacity).

A comparison of this sort is what a longitudinal study by Naigles and Tek (2017) and Naigles (2022) set out to achieve. They studied early language development in autistic children (with an average starting age of 34 months) over a period of two years, examining their grasp of both formal/syntactic and conceptual/lexical aspects of language. Based on both a review of existing literature on these issues and their own findings, they propose that 'the social difficulties of children with ASD lead the meaning-related components of their language to be relatively more impaired than the form-related components.' (Naigles and Tek, 2017, p. 1), summing up their observations in the slogan 'form is easy – meaning is hard'.

With regard to grasp of linguistic form (morpho-syntax), they report that the preschool-aged autistic children in their study were able to: (a) add appropriate plural markers to novel (nonsense) nouns, e.g., 'wug', and past tense markers to novel verbs; (b) map novel verbs in transitive frames onto causative rather than noncausative actions; (c) understand SVO word order ('the girl tickled the boy' vs. 'the boy tickled the girl'); (d) understand wh-questions; (e) understand aspectual differences (e.g., 'she's picking the flowers' vs. 'she picked the flowers'). So, apart from some delay, these children manifested no significant difference in these areas from typically developing (TD) children, despite the fact that they engage in far less talking and other communicative interactions, so their spontaneous language production is much lower than that of TD children.[21]

Moving now to these children's lexical semantic abilities (or, in my terms, their grasp of word meanings and the organization of their lexicons), this is where they seem to differ markedly from TD children. Some of the findings reported in Naigles and Tek (2017) and Naigles (2022) directly pertain to the children's lexicons, while others seem to be more a matter of the kinds of concepts they form, which is, of course, likely to impact on the nature of their word meanings. Regarding the lexicon, they report that specific word classes, e.g., mental state verbs such as 'think', 'know' and 'imagine', and words referring to emotions are significantly less present in children with autism than in TD children. This is probably not too surprising, given the well-documented autistic difficulties with social cognition (or theory of mind). A second more telling difference is that when extending a novel label for an object (e.g., 'dax') to further objects, the autistic children appeared not to have the shape bias typical of TD children (i.e., shape of an entity is typically taken to be an indicator of its kind or class, rather than color or size or texture). Some of the autistic children generalized a word's denotation on the basis of color, others required that entities have multiple properties in common. Thus, the denotations of the autistic children's words for objects are likely to be idiosyncratic, often narrower than those of TD children.

Third, their categorical induction is impaired, as compared with TD children, that is, the ability to attribute a property (e.g., 'eats grass') which has been established for one instance of a kind, say, a rabbit, to other instances of the same kind (i.e., creatures falling under a word known to the children, here 'rabbit'). This seems to be a matter of their understanding of natural kinds, which must affect the kind of encyclopedic information they incorporate in their concepts of these kinds and so may impact on their word meanings (assuming their words encode these concepts). Fourth, high-functioning autistic children were significantly less able to provide prototypical exemplars for a category word (e.g., 'bird', 'flower', 'furniture', 'game') than language-matched TD children. In short, the concepts or senses that constitute the meanings of the words they know seem to be differently organized from those of TD children. Finally, Naigles and Tek (2017) report work on young autistic adults by Perkins et al. (2006) which shows that they may use a word appropriately in a context without understanding its meaning (e.g., a young woman asked 'what does *amplifier* mean?', having just used it appropriately twice in a context);

19 They suggest that the absence of the polysemy advantage that TD children exhibit may be due to the well-known general tendency of autistic people to focus on 'specifics rather than on relationships among entities'.

20 Also highly relevant to the autistic children's difficulty in allowing words to have multiple related meanings and so to become polysemous is their widely reported literalism, that is, their tendency to give literal interpretations to language intended non-literally (e.g., metaphorically or metonymically). See Vicente and Falkum (2023) for a review and critique of explanations of this tendency, and exposition of their own view that literalism is a result of the more general autistic characteristic of strong adherence to rules and conventions.

21 One of the reviewers of the paper maintains that the evidence concerning autistic people's formal syntactic abilities is far from consistent. They note one study (of a small group of autistic adults) in which the situation seems almost the opposite to that of the autistic children surveyed by Naigles and Tek (2017), in that the participants (all of whom had very low receptive vocabularies), were unable to comprehend NP's composed of nouns and adjectives that they understood in isolation (Vicente et al., 2023). Certainly, a lot more work is needed before any general conclusions can be drawn about the formal linguistic abilities of the wide spectrum of people diagnosed as autistic.

this may be explained by the noted strength in autism of rote-learning, alongside difficulty with flexible word use.

Summing up, Naigles and Tek (2017, p. 3) say: 'when the appropriate comparisons are made, deriving meaning in a language context is shown to be disproportionately impaired in ASD, as is reflected in deficiencies in pragmatics and lexical semantics, whereas form or syntactic knowledge is shown to be either intact or proportional to other areas of functioning'.[22] However, it is not clear that the kind of evidence that Naigles and colleagues present of autistic children having a compromised or atypical lexicon is a direct reflection of their well-known social-communicative/pragmatic difficulties rather than of other aspects of their conceptual cognition, in particular, their focus on specific details and differences as reflected in their impaired category induction and atypical or absent grasp of prototypes. Thus, it would be rash to claim that the work reported so far by these researchers provides direct support for the kind of pragmatically-oriented communicational lexicon I am advocating (Section 3.3). It does, nonetheless, I think, support the more general point that the formal computational side of language and the lexical meaning/conceptual side occupy different places in the overall architecture of language, as on the narrow and broad construals of language (Section 3.2), and follow different developmental trajectories.[23]

## 4.2. Homesign (language from the ground up)

The phenomenon of Homesign provides a markedly different sort of case of a special population of communicators, one which, I believe, demonstrates in a very vivid way the pragmatic/communicational nature of the lexicon. This is the case of deaf children who are as socially attuned and interactive as typical children but who, due to their circumstances (they are born into hearing families who do not know/use any conventional sign language), have no access to a public lexicon of sense conventions. In her extensive study of their communicative development, Goldin-Meadow (2003) shows how these children, who are essentially receiving no linguistic input, spontaneously employ gestures to communicate with their non-signing families, and develop a large set of signs/words, consisting of discrete gestures paired with concepts/senses. These are negotiated and calibrated in the process of intentional communicative interactions (i.e., pragmatically), functioning initially as *ad hoc* words, whose sense has to be pragmatically inferred by the interlocutor (there are inevitably failures, leading to modifications of a gesture by the child, or, in some cases, its abandonment).

As Begby (2017) puts it: '… homesign offers a vivid illustration of the central Relevance-Theoretic claim that ostensive-inferential processes are autonomous and can serve the ends of communication even in the absence of a conventional code' (p. 699), and 'individual homesign gestures are possessed of meaning, however much that meaning fails to fall under any sort of pre-existing public norm' (p. 698), that is, these are instances of occasion-specific speaker meaning. Each of these *ad hoc* words and senses has the potential for conventionalization via frequency of use and weight of precedence, and they are not merely iconic but have a degree of arbitrariness typical of conventional word senses (Goldin-Meadow, 2003, pp. 186–87). In multiple respects, the gestures/signs produced by homesigners resemble those in established sign languages rather than the co-speech gestures of hearing people. However, it is unclear how often these do become conventionalized (and thus available for building up polysemy families) due to the exigencies of the homesigning situation: the carers' native language and the vast majority of their language use is spoken, and, even with the best will in the world, they tend not to develop much in the way of a shared lexicon with the children (Begby, 2017, pp. 707–708). This is, therefore, not a matter of cognitive limitations of the deaf child but rather of the environment in which Homesign develops, which changes dramatically if/when the child enters into a community of peers (typically in a school), as in the well-documented case of the Nicaraguan deaf children who developed a full-blown sign language over two cohorts of schooling (Senghas et al., 2005; Brentari and Coppola, 2013).

One thing is very clear: even if polysemy is lacking or scarce in Homesign, the reason for this is very different from its absence in the autistic case which, as discussed in the previous section, seems to stem from inflexibility in word use, perhaps itself due to difficulty in grasping relevant relations between meanings. The deaf homesigners, on the other hand, are highly creative in the gestures/signs they invent and modify in their drive to communicate with their caregivers. Begby (2017) gives several examples of sign usage by the homesigning child that indicate the ability to use a single sign/gesture for more than one purpose, including the following where a child is referring to her sled 'by a gesture indicating an imaginary wall space and a nail on that wall (indicated by hammering motion), this being the nail on which the sled usually hangs.' (Begby, 2017, p. 707). This is quite a complex usage (what Begby calls a 'double displacement'); simplifying somewhat, it involves the use of a homesign which means the hammering of a nail in a particular location but which is being used to refer to another object (the child's sled); in effect, this is a 'location for entity' metonymy. Whether this, in fact, became an established usage for this particular small group of interlocutors (the child and family/carers) is not clear and does not much matter – it is an *ad hoc* use of a sign, which has the potential to conventionalize and so make that sign polysemous (see also Goldin-Meadow, 2003, p. 186–188).

Furthermore, the children combine these gestures into complex structures with many of the hallmarks of typical human syntax: consistent word order, predicate frames, theta roles, hierarchical phrase structure, and recursion (Goldin-Meadow, 2003, p. 97–123). Given the complete absence of linguistic input, it is hard to envisage any explanation for this other than that a syntactic system emerges (or grows) in these children on the basis of an inbuilt language faculty, just as Chomsky has maintained.[24] While the autistic children studied by Naigles and colleagues, as discussed in the previous section, generally

22 Directly parallel results have been found in recent work testing pre-school Chinese autistic children (Su and Naigles, 2022).

23 Many thanks to Agustin Vicente for valuable help with this section, especially his detailed cautionary remarks about what implications the work reported here could (and could not) have for my position.

24 See also Carrigan and Coppola (2012), who show that even when caregivers do their best to communicate with the child in Homesign, the gesture combinations they produce lack the morphological and syntactic structure that is observed in the child's productions, so it is not the caregivers' input that is driving the child's formal linguistic development.

have good morpho-syntax, they do receive considerable linguistic input (assuming they are not deaf) and, even if they are not much interested in communication, it might reasonably be supposed that their linguistic environment is sufficient to trigger the unfolding of these formal aspects of their language. The important point for my concerns is that in both groups of children, homesigners and verbal autistic children, what we see is quite a degree of independence in the development of the formal and the meaning components of language. Linguistic form apparently came relatively easily to the autistic children but meaning was hard, leading to quite atypical lexicons, while the homesigners created meaningful words/signs as a key part of their social-cognitive drive to communicate and basic components of syntax seemed to emerge as soon as they started to combine those signs, as they do for typically developing children. While meaning (hence the communicational lexicon) is very largely dependent on social-pragmatic interaction, the organizing principles of syntax are not.[25]

## 5. Conclusion: the two parts of the language code

Decades of work within Relevance Theory has focused on the respects in which the decoded aspects of an utterance fall short of determining the speaker's meaning, with emphasis on the extensive role played by pragmatic processes of context selection, disambiguation, reference fixing, so-called 'free' enrichment (contributing to the explicature) and implicature derivation. What then is the coded part of linguistic communication? The RT answer (or assumption) has been that it is a 'semantic representation' or 'logical form', which is a conceptual structure, the conceptual part coming from the lexicon (the content of substantive words being concepts), the structural part coming from the syntax. It is typically propositionally underspecified (so a conceptual schema or blueprint) and will have multiple propositional realizations dependent on the pragmatics of different contexts of use (Sperber and Wilson, 1986/1995; Carston, 2002, 2015; Wilson and Sperber, 2004; Hall, 2008, 2009; i.a.). The central theme of the current paper is that the code has two quite distinct parts: syntax (the computational engine) and lexicon (a stored/memorized set of conventionalized phonology-meaning pairings), the one a component of the narrow language faculty, the other a component of language broadly construed, a part of the conceptual-intentional mental systems.

However, when people talk of 'the language module' (Fodor, 1983) or the linguistic decoding module (Sperber and Wilson, 1986/1995), this distinction is seldom made explicit. Decoding or parsing (language perception) is a matter of linguistic processing (performance), and certainly the relation of the syntactic parser to the system of syntactic knowledge (competence) has received a lot of attention and hard work. Still, the syntax/lexicon distinction as conceived here raises interesting issues, which, I think, have yet to be fully addressed. First, if words are the basic units of linguistic communication and comprehension (rather than roots), then word recognition is one of two basic but quite distinct processes in language comprehension, the other being the assignment of a syntactic structure to the incoming sequence of words. The question is, then, how do these two parts of the language decoder/module work together in utterance processing/comprehension? A second quite different sort of question is what does all this mean for language evolution? Again, there is a vast quantity of work on the possible continuities and discontinuities between animal communication systems and human linguistic communication. What the current picture indicates is that it makes good analytical sense to think about the advent of recursive syntax and of words separately, with the evolution of the latter to be viewed as arising in the crucible of communication and sociality more widely, while the former was more likely a result of internal changes to the thinking capacities of the human mind/brain.

25 It follows then that, while the view that all language acquisition is a product of language use in communication (the 'constructionist' position held by Tomasello (2008) and many others) seems wrong for syntax, it may well be the right way to think about the development of a communicational lexicon. See also Kissine (2021), who argues that the patterns of language acquisition and learning in autism present a strong challenge to constructionist theories of language development.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

## Funding

RC was funded by the Leverhulme Trust RF-201-078. University College London covered the publication fee (Grant No. RF-2021-078).

## Acknowledgments

Some of the material in this article was presented at the workshop on Polysemy, Oxford University, 23–24 March 2023, organized by Michelle Liu. The author is grateful to her and to all the participants for lively discussion, in particular, Nicholas Allott, John Collins, and Ingrid Falkum. Special thanks to Agustin Vicente whose detailed comments led to some significant revisions, and to the two reviewers of the paper for *Frontiers*, whose comments were encouraging and constructive.

## 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.

## Publisher's note



## References

Acquaviva, P. (2022). Word meaning: a linguistic dimension of conceptualization. *Synthese* 200:9. doi: 10.1007/s11229-022-03910-9

Bauer, L. (2018). Conversion as metonymy. *Word Struct.* 11, 175–184. doi: 10.3366/word.2018.0123

Baumgartner, L. (2022). *Pragmatic dual character classifications. Talk presented at sixth PLM conference*, Warsaw: University of Warsaw.

Baumgartner, L. (2023). forthcoming. The pragmatic view on dual character classifications.

Begby, E. (2017). Language from the ground up: a study of Homesign communication. *Erkenntnis* 82, 693–714. doi: 10.1007/s10670-016-9839-1

Berwick, R., Friederici, A., Chomsky, N., and Bolhuis, J. (2013). Evolution, brain, and the nature of language. *Trends Cogn. Sci.* 17, 89–98. doi: 10.1016/j.tics.2012.12.002

Bezuidenhout, A. (2019). "Noun-noun compounds from the perspective of relevance theory" in *Relevance, Pragmatics and Interpretation*. eds. K. Scott, B. Clark and R. Carston (Cambridge: Cambridge University Press), 174–186.

Bohn, M., Call, J., and Tomasello, M. (2015). Communication about absent entities in great apes and human infants. *Cognition* 145, 63–72. doi: 10.1016/j.cognition.2015.08.009

Borer, H. (2013a). *Taking form. Structuring sense* Oxford: Oxford University Press, *3*.

Borer, H. (2013b). "The syntactic domain of content" in *Generative linguistics and acquisition: Studies in honor of Nina M. Hyams*. eds. M. Becker, J. Grinstead and J. Rothman (Amsterdam: John Benjamins), 205–248.

Borer, H. (2014). Derived nominals and the domain of content. *Lingua* 141, 71–96. doi: 10.1016/j.lingua.2013.10.007

Borer, H. (2015). "The category of roots" in *The roots of syntax, the syntax of roots*. eds. A. Alexiadou, H. Borer and F. Schaffer (Oxford: OUP Press)

Borer, H. (2017). "The generative word" in *The Cambridge companion to Chomsky*. ed. J. McGilvray (Cambridge: Cambridge University Press), 110–133.

Bowerman, J. (2019). What's really going on with the ham sandwich?: an investigation into the nature of referential metonymy. *Int. Rev. Pragmat.* 11, 22–55. doi: 10.1163/18773109-201810012

Bowerman, J. (2021). *Referential metonymy: Cognitive bases and communicative functions*. PhD thesis, London: UCL.

Brentari, D., and Coppola, M. (2013). What sign language creation teaches us about language. *WIREs Cogn. Sci.* 4, 201–211. doi: 10.1002/wcs.1212

Carrigan, E., and Coppola, M. (2012). Mothers do not drive structure in adult Homesign systems: evidence from comprehension. Proceedings of the Annual Meeting of the Cognitive Science Society, 34. Permalink. Available at: https://escholarship.org/uc/item/87h5n840

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell.

Carston, R. (2015). "Pragmatics and semantics" in *Oxford handbook of pragmatics*. ed. Y. Huang (Oxford: Oxford University Press), 453–472.

Carston, R. (2019). "Ad hoc concepts, polysemy and the lexicon" in *Relevance, pragmatics and interpretation*. eds. K. Scott, B. Clark and R. Carston (Cambridge: Cambridge University Press), 150–162.

Carston, R. (2021). Polysemy: Pragmatics and sense conventions. *Mind Lang.* 36, 108–133. doi: 10.1111/mila.12329

Carston, R. (2022). Words: Syntactic structures and pragmatic meanings. *Synthese* 200:430. doi: 10.1007/s11229-022-03861-1

Carston, R. (2023). No unleashed expression without language. *Behav. Brain Sci.* 46:e6. doi: 10.1017/S0140525X22000802

Carston, R., and Hall, A. (2017). Contextual effects on explicature: optional pragmatics or optional syntax? *Int. Rev. Pragmat.* 9, 51–81. doi: 10.1163/18773109-00901002

Cetnarowska, B. N. D. (n.d.). Conversion as metonymy and the question of recursiveness. Available at: http://www.academia.edu/5242642/Conversion_as_metonymy_and_the_problem_of_recursiveness

Chomsky, N. (2010). "Some simple evo devo theses: how true might they be for language?" in *The evolution of human language: Biolinguistic perspectives*. eds. R. Larson, V. Deprez and H. Yamakido (Cambridge: Cambridge University Press)

Christiansen, M., and Chater, N. (2022). *The language game: How improvisation created language and changed the world*. New York: Basic Books.

Claidière, N., Scott-Phillips, T., and Sperber, D. (2014). How Darwinian is cultural evolution? *Philos. Trans. R. Soc. Lond.* 369:20130368. doi: 10.1098/rstb.2013.0368

Clark, H. (1998). "Communal lexicons" in *Context in language learning and language understanding*. eds. K. Malmkjær and J. Williams (Cambridge: Cambridge University Press), 63–87.

Clark, E., and Clark, H. (1979). When nouns surface as verbs. *Language* 55, 767–811. doi: 10.2307/412745

Dirven, R. (1999). "Conversion as conceptual metonymy of event schemata" in *Metonymy in language and thought*. eds. K. Panther and G. Radden (Amsterdam: Benjamins), 275–287.

Donnellan, K. (1966). Reference and definite descriptions. *Philos. Rev.* 75, 281–304. doi: 10.2307/2183143

Durrleman, S., Peristeri, E., and Tsimpli, I. M. (2022). The language-communication divide: evidence from bilingual children with atypical development. *Evol. Linguist. Theory* 4, 5–51. doi: 10.1075/elt.00037.dur

Falkum, I. L. (2017). The lexical pragmatics of count-mass polysemy. *Semant. Pragmat.* 10:1. doi: 10.3765/sp.10.20

Falkum, I. L. (2019). "Metaphor and metonymy in acquisition: a relevance-theoretic perspective" in *Relevance, pragmatics and interpretation*. ed. K. Scott (Cambridge: Cambridge University Press), 205–217.

Floyd, S., Jeppsen, C., and Goldberg, A. (2021). Brief report: children on the autism spectrum are challenged by complex word meanings. *J. Autism Dev. Disord.* 51, 2543–2549. doi: 10.1007/s10803-020-04687-x

Fodor, J. A. (1983). *Modularity of mind*. Cambridge Mass: MIT Press.

Goldin-Meadow, S. (2003). *The resilience of language*. New York: Psychology Press.

Grice, H. P. (1967). *The logic of conversation. Lecture 2 of William James Lectures 1967. Printed in 1975 in: Cole, P. and Morgan, J (eds.) Syntax and Semantics 3: Speech Acts. Academic Press. Reprinted in Grice, H.P.1989. Studies in the way of words*. Cambridge MA: Harvard University Press.

Hall, A. (2008). Free enrichment or hidden indexicals? *Mind Lang.* 23, 426–456. doi: 10.1111/j.1468-0017.2008.00350.x

Hall, A. (2009). Semantic compositionality and truth-conditional content. *Proc. Aristot. Soc.* 109, 353–364. doi: 10.1111/j.1467-9264.2009.00274.x

Harley, H. (2014). On the identity of roots. *Theor. Linguist.* 40, 225–276. doi: 10.1515/tl-2014-0010

Hauser, M., Chomsky, N., and Fitch, W. T. (2002). The language faculty: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Julien, M. (2007). "On the relation between morphology and syntax" in *The Oxford handbook of linguistic interfaces*. eds. G. Ramchand and C. Reiss (Oxford: Oxford University Press)

Kissine, M. (2021). Autism, constructionism, and nativism. *Language* 97, e139–e160. doi: 10.1353/lan.2021.0055

Klepousniotou, E., and Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: an advantage for polysemous but not homonymous words. *J. Neurolinguist.* 20, 1–24. doi: 10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Titone, D., and Romero, C. (2008). Making sense of word senses: the comprehension of polysemy depends on sense overlap. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1534–1543. doi: 10.1037/a0013012

Köder, F., and Falkum, I. L. (2020). Children's metonymy comprehension: evidence from eye-tracking and picture selection. *J. Pragmat.* 156, 191–205. doi: 10.1016/j.pragma.2019.07.007

Langacker, R. (1991). *Concept, Image and Symbol*. Berlin: Mouton De Gruyter.

Leslie, S.-J. (2015). Hilary Clinton is the only man in the Obama administration: dual character concepts, generics, and gender. *Anal. Philos.* 56, 111–141. doi: 10.1111/phib.12063

Levelt, W. (1999). Models of word production. *Trends Cogn. Sci.* 3, 223–232. doi: 10.1016/S1364-6613(99)01319-4

Liu, M. (2023). X-phi and the challenge from ad hoc concepts. *Synthese* 201:166. doi: 10.1007/s11229-023-04149-8

Marantz, A. (1997). *No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon*. University of Pennsylvania Working Papers in linguistics 4: Article 14.

Markson, L., and Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature* 385, 813–815. doi: 10.1038/385813a0

Naigles, L. (2022). Autism illuminates language: acquisition, development, use. *Talk given at 47th Annual Boston University Conference on Language Development (BUCLD), November 2022.'*

Naigles, L., and Tek, S. (2017). 'Form is easy, meaning is hard' revisited: (re) characterizing the strengths and weaknesses of language in children with autism spectrum disorder. *WIREs Cogn Sci* 8:e1438. doi: 10.1002/wcs.1438

Panagiotidis, P. (2014). "A minimalist approach to roots" in *Minimalism and beyond. Radicalizing the interfaces*. ed. P. Kosta (Amsterdam: John Benjamins)

Perkins, M. R., Dobbinson, S., Boucher, J., Bol, S., and Bloom, P. (2006). Lexical knowledge and lexical use in autism. *J. Autism Dev. Disord.* 36, 795–805. doi: 10.1007/s10803-006-0120-3

Pettito, L.-A. (2005). "How the brain begets language" in *The Cambridge companion to Chomsky*. ed. J. McGilvray (Cambridge: Cambridge University Press), 84–101.

Ramchand, G. (2022). "Words as things" in *For Hagit: A Celebration. Department of Linguistics, QMUL Occasional Papers in Linguistics*. eds. L. Stockall, L. Martí, D. Adger, I. Roy and S. Ouwayda

Reboul, A. (2017). *Cognition and communication in the evolution of language*. Oxford: Oxford University Press.

Recanati, F. (2002). Unarticulated constituents. *Linguist. Philos.* 25, 299–345. doi: 10.1023/A:1015267930510

Recanati, F. (2004). *Literal meaning*. Cambridge: Cambridge University Press.

Recanati, F. (2017). Contextualism and polysemy. *Dialectica* 71, 379–397. doi: 10.1111/1746-8361.12179

Scott-Phillips, T. (2015). *Speaking our minds*. Basingstoke: Palgrave Macmillan.

Senghas, R. J., Senghas, A., and Pyers, J. (2005). "The emergence of Nicaraguan sign language" in *Biology and knowledge revisited: From neurogenesis to psychogenesis*. eds. J. Langer, S. T. Parker and C. Milbraith (Mahwah, NJ: Erlbaum), 287–306.

Smith, N. V. (2019). "The source of relevance" in *Relevance, pragmatics and interpretation*. ed. K. Scott (Cambridge: Cambridge University Press), 21–26.

Sperber, D. (2000). "Metarepresentations in an evolutionary perspective" in *Metarepresentations: A multidisciplinary perspective*. ed. D. Sperber (Oxford: Oxford University Press), 117–137.

Sperber, D. (2017). Do apes produce metonymies?. Available at: http://cognitionandculture.net/blogs/dan-sperber/do-apes-produce-metonymies/

Sperber, D. (2018). Rethinking ostension: (1) a terminological issue. Available at: http://cognitionandculture.net/blogs/dan-sperber/rethinking-ostension-1/

Sperber, D., and Origgi, G. (2010). "A pragmatic perspective on the evolution of language" in *The evolution of human language: Biolinguistic perspectives*. eds. R. Larson, V. Déprez and H. Yamakido (Cambridge: Cambridge University Press), 124–132.

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

Sperber, D., and Wilson, D. (1998). "The mapping between the mental and the public lexicon" in *Language and thought*. eds. P. Carruthers and J. Boucher (Cambridge: Cambridge University Press), 184–200.

Sperber, D., and Wilson, D. (2008). "A deflationary account of metaphors" in *The Cambridge handbook of metaphor and thought*. ed. R. Gibbs (Cambridge: Cambridge University Press), 84–105.

Sperber, D., and Wilson, D. (2015). Beyond speaker's meaning. *Croatian J. Philos.* 15, 117–149.

Su, Y., and Naigles, L. R. (2022). Comprehension of core grammar in diverse samples of mandarin-acquiring preschool children with ASD. *Evol. Linguist. Theory* 4, 52–101. doi: 10.1075/elt.00038.su

Tager-Flusberg, H., Paul, R., and Lord, C. (2005). "Language and communication in autism" in *Handbook of autism and pervasive developmental disorders*. eds. D. J. Cohen and F. R. Volkman. *3rd* ed (New York, NY: John Wiley and Sons), 335–364.

Tippenhauer, N., and Saylor, M. (2019). Effects of context variability on 2-year-olds' fact and word learning. *J Exp Child Psychol.* 179, 248–259.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Vicente, A., Barbarroja, N., and Castroviejo, E. (2023). Linguistic, concept and symbolic composition in adults with minimal receptive vocabulary. *Clin. Linguist. Phon.*, 1–17. doi: 10.1080/02699206.2023.2180670

Vicente, A., and Falkum, I. L. (2023). Accounting for the preference for literal meanings in autism spectrum conditions. *Mind Lang.* 38, 119–140. doi: 10.1111/mila.12371

Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Phil. Trans. R. Soc. B* 369:20130292. doi: 10.1098/rstb.2013.0292

Wilson, D., and Carston, R. (2007). "A unitary approach to lexical pragmatics: relevance, inference and ad hoc concepts" in *Pragmatics*. ed. N. Burton-Roberts (London: Palgrave Macmillan), 230–260.

Wilson, D., and Falkum, I. L. (2015). Explaining metonymy. Talk given at the 14th meeting of the international pragmatics association (IPrA), Antwerp, Belgium.

Wilson, D., and Falkum, I. L. (2020). Understanding metonymy. Talk given at the UCL Graduate Pragmatics seminar, London, UK.

Wilson, D., and Sperber, D. (2002). Truthfulness and relevance. *Mind* 111, 583–632. doi: 10.1093/mind/111.443.583

Wilson, D., and Sperber, D. (2004). "Relevance theory" in *The handbook of pragmatics*. eds. L. R. Horn and G. Ward (Oxford: Blackwell), 607–632.

frontiers | Frontiers in Psychology

TYPE Hypothesis and Theory
PUBLISHED 31 August 2023
DOI 10.3389/fpsyg.2023.1117847



OPEN ACCESS

EDITED BY
Kate Scott,
Kingston University, United Kingdom

REVIEWED BY
Catherine Wearing,
Wellesley College, United States
Karolina Rataj,
Adam Mickiewicz University, Poland

*CORRESPONDENCE
Ira A. Noveck
✉ ira-andrew.noveck@cnrs.fr

RECEIVED 06 December 2022
ACCEPTED 01 August 2023
PUBLISHED 31 August 2023

CITATION
Noveck IA, Griffen N and Mazzarella D (2023) Taking stock of an idiom's background assumptions: an alternative relevance theoretic account.
*Front. Psychol.* 14:1117847.
doi: 10.3389/fpsyg.2023.1117847



# Taking stock of an idiom's background assumptions: an alternative relevance theoretic account

Ira A. Noveck[1]*, Nicholas Griffen[1] and Diana Mazzarella[2]

[1]Laboratoire de Linguistique Formelle, CNRS and Université de Paris-Cité, Paris, France, [2]Centre des Sciences Cognitives, Université de Neuchâtel, Neuchâtel, Switzerland

This paper begins by presenting the theoretical background of, and the accompanying psycholinguistic findings on, idiom processing. The paper then widens its lens by comparing the idiom processing literature to that of metaphor and irony. We do so partly to better understand the *idiom superiority effect*, according to which idiomatic sentences (unlike metaphoric and ironic ones) are generally processed *faster* than their literal controls; part of our motivation is to reconcile the differences between idiom processing, on the one hand, and metaphor and irony processing on the other. This ultimately leads us to Relevance Theory (RT), which has provided original insights into the processing of figurative language generally, but especially with respect to metaphor and irony. RT has paid less attention to idiomatic expressions (such as *break the ice*, *fan the flames*, or *spill the beans*), where one finds a single RT account that likens idioms to *conventional metaphors*. Through our overview, we ultimately arrive at an alternative RT account of idioms: We argue that idioms include a procedural meaning that takes into account relevant presuppositional information. For example, an idiomatic string such as *break the ice* not only asserts *initiate social contact*, it prompts the recovery of background assumptions such as *there exists a social distance that calls for relief*. This leads us (a) to apply linguistic-intuition tests of our presuppositional hypothesis, and; (b) to describe the paradigm and results from a pilot experiment. Both provide support for our claims. In doing so, we provide an original explanation for the *idiom superiority effect*.

KEYWORDS

**idioms, idiom superiority effect, figurative language processing, theoretical approaches to figurative language, relevance theory**

## 1. Introduction

An idiom is a multi-word figurative expression whose constituent parts do not readily convey its intended meaning. For example, the literal meaning of the three words in the expression *break the ice* do not in themselves reveal its idiomatic meaning, which can be paraphrased as "initiate social contact." While idioms may vary with respect to their opacity (consider how the words in *pop the question* more transparently reveal their figurative meaning, *viz.* "make a marriage proposal"), there is always a gap between an idiomatic expression's literal and figurative meaning. This ostensible gap has been at the core of theory-making and experiments on idiom processing in the psycholinguistic literature since the late 1970's, as it has been for other figures, such as metaphor, irony and metonymy (see Noveck, 2018).

The idiom processing literature has been shaped by two seminal accounts. On one side are those who argue that – despite appearances – there is nothing exceptional about idiom processing. According to the *compositional view* (Nunberg, 1978; Nunberg et al., 1994), idioms are processed word-by-word like any other expression. On the other side are those who argue that idioms call on processing that is distinctive from the sort needed to make a more literal reading. For example, according to the *direct retrieval view* (Bobrow and Bell, 1973), idioms exist in a separate, though parallel, lexicon of ordinary words. Once an independent lexicon is taken into account, one has to consider that there is an independent mode of processing. It was in the context of this debate that Swinney and Cutler (1979) defended their *lexical representation hypothesis*, which also assumed no special process for idioms; rather, these authors viewed idioms as akin to nominal compounds, such as *shrimpboat* or *hotdog.* This led the literature to consider other *hybrid views*, such as the *configuration hypothesis* (Cacciari and Tabossi, 1988), according to which people process an idiom compositionally until they recognize its idiomatic meaning, at which point it is retrieved in its entirety (Cacciari and Tabossi, 1988).[1]

Not surprisingly, this multiplicity of views has led to a large experimental literature that has provided several robust results. The best known of these is the *idiom superiority effect* which links idioms with a speed advantage when compared to non-idiomatic controls and across a wide range of tasks. Here, we mention four such cases. One comes from grammatical judgment tasks in which participants are asked to determine whether or not a sentence is a natural phrase (to appreciate a *rejectable* phrase, imagine foils such as *stranger is during*). In the context of such tasks, idiomatic expressions such as *spill the beans* (to mean *reveal a secret*) are evaluated as grammatical faster than their yoked controls, which for *spill the beans* would be something like *crate the beans* (see Swinney and Cutler, 1979). The second comes from the reading times of sentences in context. Ortony et al. (1978) reported that phrases intended to have an idiomatic reading were read more quickly than identical phrases whose context was designed to generate a literal meaning. For example, consider the idiom *let the cat out of the bag* (which also means *to reveal a secret*) when presented at the end of a vignette to describe the action of a protagonist who mistakenly revealed details of an upcoming surprise party; when presented in its control condition, this same string took longer to process when it was used to describe a saved kitten that emerged from its new owner's paper bag (we will review this one in greater detail in the Section titled "A novel approach to idiom processing"). Third, consider a paraphrase task used in an experiment from Gibbs (1980), p. 150, in which participants were presented with an idiomatic phrase, such as *he's singing a different tune*, under two conditions. In the figurative meaning condition, this phrase followed a story in which a politician changed his mind, making the phrase idiomatic; in the other, it followed a story in which a musician literally began singing a different melody. Participants were faster to endorse "he has now changed his mind" immediately after receiving the idiomatic version than they were to endorse "he's not singing the same song" after they had been exposed to the literal one. Finally, one can see speed advantages for idioms in lexical decision tasks, in which probe words (which could be a real word or not) are presented immediately after an expression; unbeknownst to the participants, real words could be either related to the idiomatic or literal meaning of strings used in a brief sentence. Cacciari and Tabossi (1988) showed that probe words related to idiomatic meanings were more quickly recognized than those linked to literal readings (also see Tabossi et al., 2009; for a review, see Espinal and Mateu, 2019). The main take-home message of these sorts of findings is that idiomatic meanings are processed as quickly or faster than their literal controls. Phenomena linked to the *idiom superiority effect* continue to inspire investigations and to be a source of discussion (for recent work, see Canal et al., 2017; Carrol and Littlemore, 2020; Mancuso et al., 2020).

Interestingly, these data on idioms appear exceptional in the scope of figurative language processing in three ways. One is that, unlike the case for metaphor and irony, idiomatic processing has been investigated in the crucible of grammatical concerns (see Fraser, 1970; Nunberg, 1978; Nunberg et al., 1994). Not surprisingly, many studies on idiom processing focus on linguistic issues, such as syntax (Gibbs et al., 1989) or on semantic decomposability (Cutting and Bock, 1997). Early studies on idiom processing treated these expressions as ambiguous between literal and figurative meanings (Cacciari and Tabossi, 1988). Gricean theory or pragmatics were rarely mentioned in these discussions (for the exceptions, see Ortony et al., 1978; Gibbs, 1980). The second way that idiom-processing research stands out is that, from early on in this literature, idioms are often investigated in isolation; this means that there is a subset of studies that investigates idioms independently of context. This too is unlike the case for other well-researched figures. Metaphors have commonly been investigated as part of a full sentence, even if it is a short one (e.g., see Glucksberg et al.'s, 1982) *Some jobs are jails*] and are just as likely to be part of a longer vignette (e.g., see Ortony et al., 1978; Gibbs, 1991; Noveck et al., 2000; Almor et al., 2007). Ironic utterances are, practically by necessity, presented as part of a rich vignette (see Jorgensen et al., 1984; Pexman, 2008; Spotorno and Noveck, 2014). Finally, the effects from idiomatic processing, on the one hand, and metaphor and irony, on the other, are mirror images of each other. While idiomatic processing findings feature how idiomatic readings tend to be faster than their controls, the findings from metaphor and irony processing indicate that these figurative readings tend to be more cognitively demanding than their controls. One of the main aims of the current work is to reconcile the findings from the idiom processing literature with those in the other figurative language processing literatures. More specifically, we aim to revisit idiom processing from a more decidedly pragmatic perspective with the further aim of addressing the *idiom superiority effect*.

In the rest of this manuscript, we take the following four steps. First, we briefly review the processing literature on metaphor and irony in order to provide a fuller picture of figurative language processing more generally. Along the way, we will describe how Relevance Theory (the post-Gricean theory that is the unifying theme of the current research topic) has been impactful to the metaphor and irony literature and specifically by (a) underlining how intention reading is central to figurative language processing and (b) showing how each figure calls for its own detailed description. Second, we go on to consider the one Relevance Theoretic (RT) account of idioms

1 There are other hybrid accounts and there appears to be disagreement about the way to classify the Configuration Hypothesis (e.g., see Titone et al., 2019). We are not concerned with these distinctions here because our goal is to provide background concerning psycholinguistic accounts on idioms before considering approaches inspired by Relevance Theory later.

that we know of (Vega-Moreno, 2001), which likens idiom comprehension to that of conventional metaphor. This section describes the added value that Vega-Moreno's account brings to the literature but then considers how it does not anticipate the *idiom superiority effect*. Third, we develop our own original account of idiom processing. Like Vega-Moreno, we root our account in RT, but we argue that idioms share characteristics with presupposition in that (at least part of) their processing involves satisfying an idiom's assumed preconditions. Once this assumption is made, one is in a better position to account for the *idiom superiority effect*. Finally, we present an armchair linguistic test that aims to test our account as well as an experimental paradigm that served as the basis for a pilot study. The conclusions from both tests provide support for our account.

## 2. Figurative language processing: looking under the hood

### 2.1. A focus on metaphor and irony processing

The basis for many of the early studies on figurative language was inspired by Gricean Theory, which, as Gibbs and Colston (2012, p. 65) wrote, "assumed that people must always do a complete literal analysis of an expression before any pragmatic information is evaluated to derive speaker meaning, which in turn implies that figurative language processing must always take longer than literal speech comprehension." As data accrued on figurative language processing, it became clear that this need not be the case. At some point, Grice's so-called Standard Pragmatic Model (fairly or unfairly) became the reference that experimental post-Gricean accounts typically argued against (for a discussion, see Noveck, 2018). The upshot is that several new accounts emerged, two of which played a large part in forming the psycholinguistic literature on figurative language processing.[2] One is Gibbs's (1986) and Gibbs et al.'s (1989) *Direct Access* view, which claims that the non-literal meaning of a given word or phrase is accessed without considering its literal meaning. However, as even more data were collected, much of them showing that metaphoric and ironic meanings often *are* associated with slowdowns compared to literal readings, it was hard to maintain this argument [see (Gibbs and Colston, 2012), for arguments *against* the Direct Access view]. The other important position to emerge on figurative processing was championed by Giora (1997), who argued that conventional and frequent (i.e., the most salient) meanings of words are processed first. Thus, slowdowns related to figurative readings are likely due to those situations in which participants consider literal meanings first. However, according to Giora's *Graded Salience* view, literal meanings *can be* superseded by accessing a word's figurative meaning (with respect to irony, see Giora et al., 1998, 2007; Giora and Fein, 1999; Schwoebel et al., 2000).

As this theoretical introduction indicates, one of the main dependent measures in the study of metaphoric and ironic processing is the relative reading time speeds comparing a figurative reading of a sentence to a baseline. If one were to find evidence indicating that the understanding of a metaphoric or an ironic sentence, say, generates longer latencies than its literal reading, one could argue that figurative processing is effortful; if one does not find differences (i.e., null results), one could argue that there is nothing unique about figurative language. The juxtaposition of these two kinds of possible findings characterizes exchanges and debates in the psycholinguistic literature since its beginnings with respect to ironic and metaphoric materials. It would be fair to say that, overall, one finds that figurative readings *can* indeed be accessed as fast as literal ones; however, *ceteris paribus* and with minimal context, figurative language processing is generally associated with slowdowns.

To appreciate the kind of data that these theories aimed to account for, one can hark back to some of the earliest studies on metaphor processing. For example, Ortony et al. (1978) prepared vignettes in which a test sentence could be placed in one of two contexts, one that would render the test sentence metaphorical and the other literal. Consider the test sentence, *Regardless of the danger, the troops marched on* when it is preceded by just one line of text. In order to elicit a metaphoric reading, that preceding line concerned children who were annoying their babysitter; in order to elicit a literal reading, the authors provided a line concerning soldiers in battle. With such limited contexts, the test sentence was read significantly more slowly in the metaphoric condition than in the literal condition. However, when the preceding text was expanded to a paragraph (thus providing much more referential detail), the latencies of figurative versus literal readings of critical test sentences were dramatically reduced (for another early study in the same direction with short sentences and ERPs as dependent measures, see Pynte et al., 1996). For metaphor (as well as for irony), it is generally the case that figurative meanings take longer to process than their literal controls, and especially when there is little background information. However, contextual or experimental features can indeed reduce these differences, to the point that figurative cases can appear as fast as their literal controls. Note, again, that the *idiom superiority effect* makes a different stronger claim, which is that idiomatic readings are processed *faster* than their literal controls.

### 2.2. Relevance theoretic approaches to irony and metaphor

Relevance Theory has played a prominent role in accounting for irony and metaphor. According to Sperber and Wilson (1986/1995) determining the meaning of an utterance is part of a listener's effort to understand the speaker's intended meaning which is always inferred (even when it consists in deriving a literal interpretation of an utterance). The inferences involved, however, make the comprehension of an utterance vary with respect to the effort they require. Both the sentence meaning and its context contribute to making some interpretations more easily derivable than others. For both irony and

2 There are multiple post-Gricean approaches, such as the *Constraint Satisfaction* view (Pexman et al., 2000; Pexman, 2008), which states that multiple factors can influence processing of irony in parallel, including the speaker's reputation and role, that can result in equivalent speeds for comprehending ironic sentences and their literal counterparts (Ivanko and Pexman, 2003; Katz et al., 2004). We do not aim to summarize all current accounts of figurative processing here, but to provide the academic experimental context in which Relevance Theory operated before turning to our approach to idiom processing.

metaphor, it is not a single application of a general rule (such as *Direct Access* or *Graded Salience*) that generally determines how figurative language is processed and whether a figurative meaning requires more (or less) effort to process than its literal use. Rather, comprehension difficulty depends on the intention that the speaker wants to communicate and the inference-making that it takes to read that intention. While there were Relevance Theorists who carried out experiments from early on (on irony, see Jorgensen et al., 1984), nearly all of their arguments about figurative language were theoretical until the turn of the century when experimental studies inspired by Relevance Theory became more commonplace (see Noveck et al., 2000; Noveck and Sperber, 2007).

As far as irony is concerned, RT emphasizes the *echoic* use of language, in which comprehension depends on detecting that the speaker wants to convey a skeptical, mocking or dissociative attitude about a previous thought. For example, imagine two professional opera performers who unexpectedly sing horribly (see Spotorno and Noveck, 2014). When one singer says (1) ironically to the other, she is mocking herself and her colleague for having had more lofty expectations.

> (1) Tonight we gave a superb performance.

Spotorno and Noveck (2014) showed that ironic readings of lines of text like those in (1) do indeed take longer to process compared to literal readings of the same lines (to appreciate [1] literally, imagine a context in which the vignette describes the singers as having performed well before the critical test sentence is presented). However, the authors also reported *Early-Late* (trial) *effects*, i.e., differences in reading time speeds between the two conditions largely disappear by the time a participant comes to the end of their experimental session (particularly when ironic utterances in a reading task arise consistently after a negative event). Their explanation was that it is intention reading (or Theory of Mind) that intervenes as the speaker's – or perhaps the narrator's – intention becomes more obvious as the number of ironic reactions populate the experimental session (for a recent extension on such intention-reading claims, see Ronderos et al., 2023). Incidentally, neuro-imaging studies show that participants' irony-processing appears to rely on brain regions highly associated with Theory of Mind processes (Spotorno et al., 2012). In short, for the RT approach, it would not be surprising to find that ironic readings could be carried out with a speed that is comparable to literal readings as long as the ironic line is consistent with the speaker's apparent intention.

In contrast, the standard RT account views metaphor as a form of 'loose use of language' comparable to other phenomena, such as hyperbole and neologism, in which the meaning communicated by the use of a word in context ultimately differs from the linguistically encoded meaning of that word. Through a general pragmatic process known as 'concept adjustment' (Carston, 2002; Wilson, 2003), a word could convey a more specific concept, as the word *drink* does in *Let me buy you a drink* when heard at a bar to mean "an alcoholic drink." Or, it could convey a concept that is more general than the lexical concept. For example, the shape in *France is hexagonal* is very much a loose use of that mathematical object. Metaphors, according to RT, involve a combination of narrowing and broadening (Carston, 2002) that helps guide the listener to understand the concepts that the speaker intends to communicate.

Rubio-Fernandez (2007) investigated metaphor in a Relevance framework through a cross-modal priming study. She argued that the enhancement of the relevant properties of the metaphor, and suppression of those that are irrelevant for the figurative interpretation, is a necessary process in metaphor comprehension. In order to test her claim, she presented 20 two-sentence-long vignettes whose second sentence would often conclude with a metaphor. For example, participants would read sentences such as (2):

> (2) Nobody wanted to run against John at school. John was a cheetah.

By presenting probe words immediately after the metaphor (i.e., zero seconds after the final word) or else at 400 ms or 1,000 afterward, she was able to profile the way metaphoric meanings emerge. That is, she presented three different kinds of probe words – (i) an unrelated term (e.g., *plant*), (ii) a superordinate, literally related, term (*cat*) or; (iii) a distinctive, figuratively related, term (*fast*) – after items like those in (2) and reported three findings. First, the *immediate* reactions to the three types of probes revealed that the unrelated meaning is significantly slower than both the metaphor's literal superordinate meaning *and* the metaphor's intended meaning. Second, listeners continued to show a speed advantage for the superordinate probe over the unrelated one at 400 ms but this preference disappeared at 1,000 ms. Third, like the superordinate probe, the distinctive-property probe also had a speed advantage over the unrelated one at 400 ms; unlike the superordinate word, the distinctive probe maintained its advantage over the unrelated one at 1000 ms. She concluded that metaphor interpretation involves enhancing relevant properties of the metaphor vehicle while actively *suppressing* the superordinate associations.

## 3. A relevance theoretic approach to idioms: are they akin to conventional metaphors?

The question of how idioms are processed has received less attention from relevance theorists as compared to other figurative uses of language. One notable exception is the work from Rosa Vega-Moreno (2001, 2003, 2005), whose relevance-theoretic account views idiom comprehension as comparable to other loose uses of language. Her account underlines the importance of thinking of idiomatic expressions as essentially irreducible to their literal paraphrase; as she puts it (Vega-Moreno, 2001, p. 76): "idioms cannot be paraphrased without loss." For instance, she argues that the idiom *kick the bucket* cannot be aptly paraphrased with the verb *die*. Although the two encode conceptual representations that are logically related (anybody who kicked the bucket died), the mental representation associated with the idiomatic expression contains additional information about the manner of death (people who kicked the bucket presumably died suddenly and unexpectedly), the attitude that one has (for this example, we assume she means the attitude toward the deceased) as well as "something imagistic" (Vega-Moreno, 2001; p. 76), all of which are not associated with the concept encoded by the verb *die*. As a result, far from being a rhetorical device, the use of an idiomatic expression would be motivated by the speaker's intention to convey a

meaning that could not be conveyed otherwise. But how is this meaning inferred by the hearer?

Vega-Moreno (2003, 2005) suggests that the interpretation of an idiomatic expression relies on a pragmatic inferential process that can take as its starting point the meaning encoded by the idiom string (a holistic and structured conceptual representation), as well as the lexical meaning of its constituents. When processing the idiomatic string, both the structured phrasal concept encoded by the idiom and the concepts lexically encoded by the individual constituents are activated. As with metaphors or other loose uses of language, these meanings are subject to a process of conceptual adjustment, which allows the construction of *ad hoc* concepts. As a result, they can convey an occasion-specific meaning related to the particular use of the word or phrase at issue and can ground the derivation of relevant implications (see Wilson and Carston, 2007).

According to Vega-Moreno (2003), the interpretation of idiomatic expressions typically involves the interplay between the pragmatic adjustment of the concepts lexically encoded by the idiom's constituents as well the construction of an *ad hoc* phrasal concept. To illustrate this, consider her main example:

> (3) I cannot stand the way my boyfriend is tied to his mother's apron strings.

To interpret the expression "tied to his mother's apron strings," a listener could start with the encoded concept TIE and pragmatically broaden it to include in its denotation any process in which some degree of attachment is involved (thus constructing an *ad hoc* concept TIE*). Crucially, though, this on-line conceptual adjustment would be complemented by accessing the meaning of the idiomatic expression as a whole:

> At some point during this process, the concept encoded by the idiom string as a whole is retrieved from memory ([TO BE TIED [TO [ONE'S MOTHER'S APRON STRINGS]]]*). Rather than involving a switch of processing mode, the hearer takes this concept also as a further clue to the speaker's meaning and he starts considering some of its accompanying information (e.g. the assumption that someone with this property is too close to their mother, not independent enough for their age, and so on) (Vega-Moreno, 2003, p. 313)

As a result, idiom comprehension would involve a process of conceptual adjustment of the meaning of the individual constituents (which typically involves broadening) as well as a process of conceptual adjustment of the stable conceptual representation associated with the idiomatic string in memory (which typically involves narrowing). Indeed, because the conceptual representation that is associated with the idiom string in memory is often unspecified, it regularly requires some pragmatic specification.

Crucially, the relative role of conceptual adjustment with respect to the meaning of individual constituents depends on the degree of decomposability and transparency of the idiomatic expression; that is, each constituent contributes independently and in an identifiable way to the idiomatic interpretation. In the case of decomposable idioms, the greater the idiom transparency, the greater the contribution of the pragmatic adjustment of the meaning of individual constituents. In the case of non-decomposable idioms (e.g., *chew the fat*), the meaning of the constituent words does not contribute at all to the recovery of the idiom meaning, so any process of conceptual adjustment at a word level may disrupt, rather than contribute to, the derivation of the idiomatic interpretation. Although consistent with some available empirical data suggesting that the understanding of compositional idioms is often facilitated as compared to non-compositional strings (Gibbs, 1991), note that this account does not address the *idiom superiority effect*. At least it is not clear what would make this explanation account for faster idiomatic reading times when compared to its controls.

Vega-Moreno's account displays, however, some interesting features, which we incorporate in our analysis of idiom comprehension. First, it acknowledges the richness of the meaning that is conveyed by idiomatic expressions and its irreducibility to a literal paraphrase. Second, it emphasizes the pragmatic dimension of idiom understanding: far from being a matter of pure linguistic decoding, the processing of idioms involves a great deal of pragmatic inference to recover the speaker's intended meaning. Indeed, in line with the relevance-theoretic framework, Vega-Moreno (2003) conceives conceptual adjustments at the word or phrasal levels as part and parcel of the search for a relevant interpretation of the speaker's utterance containing the idiom. As a result, idiom comprehension relies on pragmatic enrichments typically involving broadening of the lexically encoded meaning of individual constituents as well as narrowing of the conceptual representation associated with the idiomatic string as a whole. This involves "a simultaneous adjustment of word, phrase and sentence meaning which take[s] place during the process of deriving the explicit content, *context* and implicatures" (Vega-Moreno, 2003, p. 312, our emphasis).

In what follows we take these points a step further and suggest that a full-fledged account of idioms requires spelling out how the idiomatic interpretation contributes to the derivation of the appropriate context. We argue that understanding an idiomatic expression involves the derivation of a series of background contextual assumptions, whose use is critical in order for the idiom to be understood.

## 4. A novel approach to idiom processing

### 4.1. The background

The genesis of the account that we are about to present emerged while reviewing experimental papers on idiom processing in the context of the figurative language processing literature generally. It was observed that – typically – when an idiomatic expression was employed as part of a vignette in a behavioral task, the figurative meaning was sensible because the context contained several elements that justified its use, i.e., it was felicitous in context. For example, consider the item below (4a) from one of the early studies on idiom comprehension (Ortony et al., 1978):

> (4a) Dean spoiled the surprise that Joan had been planning for their mother's birthday party. When he realized what he'd done, he apologized for having let the cat out of the bag.

One can see that the reader is informed that there was a surprise that was ruined so the expression *letting the cat out of the bag* fits with

a prior situation. In contrast, when an idiomatic phrase is used to convey a literal meaning, it would naturally make the idiomatic meaning nonsensical in context, if one were to assume that the figurative meaning was indeed generated. For example, consider the literal control (4b) for the vignette above:

> (4b) Walking back from the store, Anne found a kitten which she put in with her groceries. She got home and her puppy went wild when she let the cat out of the bag.

In this case, *letting the cat out of the bag* is designed to be understood literally. Note though that – if the idiomatic meaning were to be generated – it would come without any contextual support. There is no secret that was betrayed so the idiomatic reading would be incongruous with the previous information.

This appears to be a general feature of idiom studies that use vignettes: when a string is employed and intended to be understood idiomatically, its specific preconditions were met in the prior context and when the literal controls of idioms are employed, (a) the investigated strings are assumed to be stripped of their idiomatic meaning and; (b) the vignettes are naturally presented without the contextual support that would make the string's idiomatic meaning felicitous [for other such examples, see (Gibbs, 1980)]. With this insight in tow, one has the beginnings of a sensible explanation for the *idiom superiority effect*. That is, it indicates that the well-known effect is not necessarily due to idiom processing being in some way accelerated; it is arguably due to the fact that the literal control items (which still use idiomatic strings) prompt slowdowns because they (a) likely generate figurative meanings which are then (b) incongruous with the context.

The upshot worth noticing is that the felicitous use of idiomatic expressions appears to require some contextual preconditions (which vary from idiom to idiom) that are needed to make the idioms apt. For instance, the idiomatic meaning of *break the ice* is felicitous in contexts characterized by an initial tension between strangers (see Levorato et al., 2004), while *fan the flames* is sensible in contexts where there is a pre-existing and ongoing conflict and *spill the beans* is reasonable in contexts in which there is a secret to reveal, and so on and so forth. We take this observation as a starting point to make a fundamental suggestion: that idioms are accompanied by a set of background assumptions, which verge on being presuppositional. To develop this proposal, we first clarify the notion of presupposition that we have in mind (section "Presuppositions") and then elaborate on the way in which idiomatic expressions can also be conceived as carrying presuppositional-like effects (section "Idioms and presupposition-like effects").

## 4.2. Presuppositions

The idea that information can be presupposed, as opposed to being explicitly asserted, has a long history in philosophy of language and linguistics (for an overview, see Beaver et al., 2021). Presuppositions are standardly described as backgrounded information, that is, information that is old, given or taken for granted by the interlocutors (or at least presented as such). For the purpose of this paper, we focus on the way in which this notion has been integrated into Relevance Theory. In her seminal work, Simons (2005) describes presuppositions as "relevance requirements" or "relevance establishers"; that is, they are background assumptions that contribute to the relevance of the overall interpretation by giving access to a context in which further assumptions (the explicatures or implicatures of the utterance) can achieve contextual effects (see also de Saussure, 2013; Mazzarella and Domaneschi, 2018). According to Simons, presuppositions are thus "the propositions which that addressee must accept for the utterance to be relevant in the way intended by the speaker" (Simons, 2005, p. 333). For instance, to achieve relevance in the way intended by the speaker, an utterance, such as *Even Trump admitted that climate change is real* (adapted for our purposes from Simons) requires the background assumption that Trump is a particularly unlikely person to make such an admission, based on his previously shared views on the matter. The addressee needs to accept this presupposition to infer some intended implications, such that the evidence of rapid climate change is undeniable, that this should be a concern for the environmental policy of all parties, etc.

The linguistic literature on presupposition has identified a variety of lexical items or constructions that trigger the derivation of presuppositions (and are thus known as "presupposition triggers," see Karttunen, 1974; Levinson, 1983). These include expressions such as factive verbs or change-of-state verbs, which trigger presuppositions that are undetachable from what is said. For instance, by uttering "Deirdre *left* the house," the speaker presupposes that she was in the house immediately before the reference time and asserts that she is no longer in the house (where the asserted content cannot be expressed without triggering the presupposition). Furthermore, the class of presupposition triggers also include lexical items such as *again*, *too*, *even*, whose only function seems to be the triggering of the presupposition itself. By uttering "Deirdre laughed *again*," the speaker is using *again* to presuppose that she laughed before (for a discussion on the distinctive features of these "dedicated presupposition triggers," see Simons, 2005). Drawing on the relevance-theoretic notion of "procedural meaning" (Blakemore, 1987, 2002; Simons, 2005) suggests that presupposition triggers encode dedicated procedures, which guide the inferential comprehension process by imposing constraints on the construction of contexts. These expressions can thus "indicate that the speaker intends (the truth conditional content of) her utterance to be interpreted relative to a context which contains 'the presupposition'" (2005, p. 349).

The role of context has been widely investigated with respect to the processing of statements containing presupposition triggers. Indeed, many experimental studies have explored the contrast between situations in which a presupposition is contextually supported, or "satisfied," and a situation characterized by a "presupposition failure" which cannot be readily repaired (e.g., see Ferretti et al., 2008, 2013; Jouravlev et al., 2016; Shetreet et al., 2019; for a review, see Schwarz, 2015 or Reinecke, 2020; for formal distinctions, see Glanzberg, 2003). In the former case, the linguistic context already includes or entails the background assumption that is linguistically triggered by the presuppositional statement. In the latter case, not only the presupposed content is unavailable (not taken for granted by the interlocutors), but it is also inconsistent with the immediately preceding linguistic context, thus making it impossible for the hearer to accept it (or "accommodate" it, see Lewis, 1979). Consider a straightforward example from Jouravlev et al. (2016), who compared two kinds of stimuli (e.g., 5a and 5b) as part of an EEG study:

> (5a) Jake had tipped a maid at the hotel once before. Today he tipped a maid at the hotel again.
>
> (5b) Jake had never tipped a maid at the hotel before. Today he tipped a maid at the hotel again.

These authors reported late widely distributed positivity after the onset of the trigger *again* (indicative of an early arriving P600 effect) when the presuppositional trigger was inconsistent with the previous content (in 5b) as opposed to consistent with the previously stated context (in 5a). This indicates that presupposed information is processed differently depending on whether the context is supportive or not. Not surprisingly, presuppositional content is integrated more smoothly when it is consistent with preceding information.

## 4.3. Idioms and presupposition-like effects

We want to suggest that idioms are also typically accompanied by a set of background assumptions, making these figures verge on being presuppositional; i.e., the relevance of an utterance containing the idiomatic expression depends on the recovery of these assumptions. For instance, by using the idiom *break the ice* in "Joey broke the ice by making a joke," the speaker appears to presuppose a range of background assumptions that includes the presence of an initial tension among a relevant group of people, a tension which is palpable and which nobody had yet tried to mitigate. We adopt Simons's (2005) language to describe these background assumptions as presuppositional. That is, they are meant to contribute to the relevance of the overall interpretation by setting up a context in which the claim "Joey broke the ice by making a joke" can be interpreted as implying that Joey was motivated to change this uncomfortable interpersonal dynamic (thus functioning as "relevance establishers"). Similarly, the idiom *fan the flames* in "She fanned the flames" appears to presuppose the existence of preexisting tension, characterized by a certain degree of animosity among the people involved. Recovering these background assumptions plays a crucial role in constructing the context in which "She fanned the flames" can be interpreted as suggesting that she acted in a way that is likely to feed this existing conflict and aggravate it, to worsen the personal relationship at stake, etc. Similar considerations can be applied to a variety of idioms (*make a killing*, *hit a wall*, *clip one's wings*, *spill the beans*, etc.), thus indicating that processing an idiom routinely involves accessing a variety of background assumptions that shape the context of interpretation.

How are these background assumptions brought about in the interpretation process? In the literature on presuppositions, we find a well-established distinction between the so-called "pragmatic presupposition" and "semantic presupposition" (see, e.g., Potts, 2015). While the former is entirely pragmatically motivated (see also de Saussure, 2013 on "discursive presupposition"), the latter traces to conventional aspects of the meanings of specific words and constructions, the class of presupposition triggers discussed above. The status of the background assumptions associated with an idiomatic expression is far from being settled, and it may well differ from idiom to idiom. In what follows, though, we explore our original hypothesis, which is that idioms encode procedural meanings that work as instructions for the recovery of these assumptions.

According to our hypothesis, idioms encode not only a conceptual component but also a procedural one. Following Simons' account of presupposition triggers as procedural meanings, idioms are thought of as encoding procedures to construct the relevant context of interpretation, one which includes the set of background assumptions which make the use of the idiom felicitous. For instance, the idiomatic expression *break the ice* would thus encode procedural indications to recover assumptions related to the presence of palpable tension among the people at issue. As suggested by Wilson (2011, p. 9), "[t]o say that a word encodes a certain concept or procedure is to say that the linguistic system is linked to the rest of the cognitive system in such a way that activating the word will systematically activate the associated concept or procedure." It follows from this that if idioms encode both a phrasal concept and a procedure, processing the idiomatic string will systematically result in activating both, thus triggering a process of inferential reconstruction of a set of relevant background assumptions.

In line with standard examples of presupposition triggers, most idioms would typically prompt the recovery of a precise and identifiable set of background assumptions. It is also worth noticing that certain idioms additionally invite the recovery of broader (and vaguer) arrays of assumptions, attitudinal dispositions or imagistic components, that could be described in terms of so-called "non-propositional effects" (Wilson and Carston, 2019). Consider, for instance, how *kick the bucket* conveys a specific attitude toward the deceased. That is, reconsider Vega-Moreno's (2003) example for *kick the bucket* in (6):

> (6) Has horrible old Mr. Thomas kicked the bucket yet?

Clearly, the dissociated, distant, or even negative attitude toward the referent of this idiom (*horrible old Mr. Thomas*) plays a role in comprehending the idiom. To appreciate the role of attitude, compare (6) to *Has the love of my life, my inspiration, Tom, kicked the bucket yet?* In this latter case, the choice of idiom, taken at its face, appears to be incongruous.

We suggest that the hypothesis that idioms encode procedural meanings meshes well with these observations. Interestingly for our purposes, the notion of procedural meaning has also been employed to capture the expressive dimension of a range of communicative devices – interjections, emotional prosody, expletives, etc. – which are regularly associated with the expression of an emotive attitude (see, e.g., Wharton, 2003; Wilson and Wharton, 2006; Blakemore, 2011). In all these cases, the encoded procedures are taken to activate representations of emotional states, evaluative contents or attitudinal descriptions (for a discussion, see Carston, 2016). The notion of procedural meaning thus seems to be well suited to account for the presuppositional-like effects of idiomatic expressions, even when these pertain to less determinate and more nuanced contents.

Finally, this hypothesis can shed additional light on the claim that idioms are irreducible to their literal paraphrase, so that - as discussed before – *kick the bucket* cannot be paraphrased as *die* without loss (Vega-Moreno, 2001). One possibility is to think of the meaning that is lost in the paraphrase as linked to the procedural meaning idioms encode. It is arguably the procedural meaning encoded by the idiom that allows for the richness of the content inferred via the idiomatic strings when compared to a literal, less nuanced, paraphrase.

## 5. Testing our presuppositional claims

Given the prominent place that empirical data has played in developing our own hypothesis, it is only appropriate that we employ tests to evaluate our original claims. We go about this in two steps. One is to employ a well-known empirical test that relies on our linguistic intuitions and the other is a more severe experimental pilot that collects psychological measures. This is what we turn to in the next two sections.

### 5.1. Armchair observations

When one hears the utterance *Noemi stopped smoking*, it implies that she smoked in the past (this is *presupposed* content) and that she currently does not smoke (this is often referred to as *at issue* content). One of the standard linguistic-intuition-based tests of presupposition aims to determine whether *presupposed* content projects across a specific range of grammatical contexts, even as these contexts modify the *at issue* content. For example, a presupposition expressed under negation would maintain the existence of presupposed content even as the at-issue content is reversed: Upon hearing *it is not the case that Noemi stopped smoking*, the presupposed content (that she smoked in the past) is maintained, but the at-issue information is reversed, i.e., one would infer that Noemi currently smokes. Thus, the presupposition is said to project. Simons (2005) refers to these projection tests in specific grammatical contexts as Basic Projection Facts, which we list in (7) below (also see Langendoen and Savin, 1971; Chierchia and McConnell-Ginet, 1990):

> (7) Given a sentence S which, when uttered, is typically understood to presuppose p,
>
> utterances of a sentence S′ will typically also be understood to presuppose p, where:
>
> a. S′ is the negation of S.
>
> b. S′ is the yes/no question formed from S.
>
> c. S′ is a conditional with S as its antecedent.
>
> d. S′ embeds S under an epistemic modal.

One immediate way to test our idiom-related claims then is to determine whether given presuppositional information that we claim is tied to idioms projects in these grammatical environments in a way similar to the classic presuppositional cases. That is, if the presupposed content appears to remain in the classic test environments, even as the at-issue information might not appear to, this would provide some intuitive support to our claim. Let us consider *Joey broke the ice by making a joke*, which will be adopted later into our experimental task. As part of our armchair test, we transform this phrase and create four new ones that are distributed across the four environments described above. These are expressed as (8a-d):

> (8) *Example*: Joey *broke the ice* by making a joke.
>
> a. *Negation*: Joey did not *break the ice* by making a joke.
>
> b. *Question*: Did Joey *break the ice* by making a joke?
>
> c. *Antecedent of conditional*: If Joey were to *break the ice* by making a joke, it would have no positive effects.
>
> d. *Possibility modal*: Joey might *break the ice* by making a joke.

Again, the question is whether the presupposed content for this particular idiom in this particular sentence – that there was some pre-existing social tension (before an effort was made to relieve that tension) – persists across these grammatical environments. In our reckoning, they do. In (8a), under negation, the asserted content has been negated (the social tension was not relieved or the joke did not succeed) but the presupposed content remains (that there was some pre-existing social tension). In (8b), whether one responds affirmatively or negatively, the felicity of the answer depends on assuming that there was some contextual reason that called on breaking the ice. Likewise for the remaining cases: whether or not the "relieving social tension" meaning is confirmed, reference to presupposed content (reference to the existence of some prior social tension) remains. It appears then that our claim passes its initial test. While we do not want to get ahead of ourselves, it is noteworthy that even if our findings apply to only a subset of idioms, this amounts to a novel characterization of idioms.

### 5.2. Testing our claims experimentally: initial findings

Crucially, by appreciating the presuppositional effects of idiomatic expressions, we can better understand why the use of a given idiom is felicitous only under certain circumstances. To illustrate this, consider the following two examples:

> (9) a. Joey was enrolled in a competitive biology course. At the beginning of the semester, no one dared to speak to each other. Therefore, Joey *broke the ice* by making a joke.
>
> b. Joey was enrolled in a competitive biology course. By the end of the semester, everybody in the class got to know each other. Therefore, Joey *broke the ice* by making a joke.

In (9a), *broke the ice* is used felicitously because the array of background assumptions triggered by the idiom (i.e., that there exists social tension among the classmates) is consistent with the assumptions provided by the preceding statement ("no one dared to speak to each other"). In contrast with this, in (9b), the array of background assumptions that the speaker appears to presuppose by the use of *broke the ice* is inconsistent with her preceding statement ("everybody in the class got to know each other"), thus generating the perception of an infelicitous discourse continuation. The contrast between (9a) and (9b) is thus comparable to the contrast discussed in section "Armchair observations" above between cases of satisfied, or contextually supported, presupposition and cases of "presupposition failure" [see our examples in (5a) and (5b)].

In order to more severely test our hypothesis, the first and second authors prepared a pilot experiment (see Griffen and Noveck, 2023) which was based on the insight described above, i.e., that idioms concern not only at-issue information (such as *initiated social contact* for *broke the ice*) but presupposed information as well (*that there was some pre-existing social tension*). In this way, idioms are similar to presuppositions. This led us to develop a paradigm (inspired by a study on presuppositions from Shetreet et al., 2019), in which participants would receive brief vignettes.

In this pilot, 18 idioms were investigated and, for each, a precondition was readily identified. As before, let us provide a couple of other examples. For "bury the hatchet," which can be loosely translated to mean "make peace," it presupposes that there was discord previously. For "spill the beans," which (as described earlier) means something akin to "reveal a secret," it necessitates that there was a secret to be kept. When such a precondition is satisfied in the discourse, one would expect the idiom's use to be felicitous and facilitated; when the precondition is not met (or not fully met), it would make the idiom's use appear infelicitous (or, at least hard to accommodate).

This led us to prepare vignettes for the purposes of a self-paced reading study. Each vignette comes in one of two varieties, one that will support the idiom's presuppositional content (as in 9a) and one that does not (as in 9b) above (these are combined and reprised together in (10) as we highlight the task's experimental features). To provide a little variety we also show our vignette for *spill the beans* in (11). Both (10) and (11) underline three experimental features. One is that the slashes indicate the point at which a participant advances the text so there are always five such reading segments. A second is that there were two kinds of second sentences (or second segments), as first shown across (9a) and (9b): one type of second-sentence will later make the idiomatic string felicitous and the other type (in brackets) will make it infelicitous. The third is that the last three segments (the third through fifth) make up the third sentence of the vignette, among which one finds the idiomatic string always occurring in the fourth segment.

> (10) Joey was enrolled in a competitive biology course./At the beginning of the semester, no one dared to speak to each other. [By the end of the semester, everybody in the class got to know each other.]/Therefore, Joey/ *broke the ice*/by making a joke.

> (11) Nick is organizing a huge birthday party for his mother next week at their house./She knows nothing about the party because it is planned as a surprise [She has taken over the planning because she loves entertaining.]/Last night, Nick/*spilled the beans*/about the event.

Twenty-four vignettes were presented as part of our self-paced reading task (there were also numerous filler items that had nothing to do with presuppositions or idioms). Eighteen items were devoted to idioms and were distributed randomly across three conditions. Participants received a story context (like the ones in [10] and [11]) that led to an idiom that was (i) felicitous, (ii) infelicitous, or else (iii) a control in which the felicitous context was presented but included an invented nonsense idiom instead. To make this concrete, the control version of (iii) for (11) kept the "felicitous" second sentence above, but replaced *spilled the beans* with *cramped the air*. The 18 idiom-potential vignettes were rotated so that every story context was presented as the source of one of the three conditions and so that every participant received one of the three. All told, an individual participant received six randomly chosen idiomatic strings that were presented in a felicitous context, six whose contexts were infelicitous with respect to the idiom's presupposed information (and thus required some accommodation), and six that used nonsense idioms (where another idiom would be appropriate). The remaining six items were control items drawn from Shetreet et al.'s EEG study Shetreet et al. (2019) on presuppositions, which recorded reactions to (a specific word in) a sentence that was (a) downstream from a factive presupposition and; (b) that made the sentence either consistent or inconsistent with the prior context. For an example, in the item in (12) below, participants received a second sentence that would make the third sentence appear consistent [or inconsistent] with the previous context:

> (12) Bruce taught a class on quantum physics./He **saw** that his students had *mastered* [were *confused* by] the material. /Almost all of them/scored perfectly/on every test.

Items such as these were included as a sanity test. It was assumed that we would extend Shetreet et al. (2019) results (which concerned ERPs to the underlined term in the third sentence) by finding reading time slowdowns for those cases where the third sentence is inconsistent with the presupposition in its context.

Our online participants progressed through each story by pressing the spacebar on their keyboard and would occasionally receive a comprehension question. Importantly, the idiom string, which appeared in the fourth segment, as well as the final segment which appeared in the fifth, ultimately provided dependent variables. Our expectation was that there would be significantly faster reading times for the fifth segment when it appears after a felicitous idiomatic string rather than after an infelicitous one. In other words, we expected slowdowns when the idiomatic string was used in non-felicitous context; likewise, we expected our nonsense idioms to produce slowdowns, too.

To be brief, we can say that our results aligned with our predictions. To provide a little detail, Griffen and Noveck (2023) point out three findings concerning the reading times of the last (the fifth) segment. First of all, the findings extend (Shetreet et al's, 2019) outcomes with reading times, which further validates their paradigm and provides a benchmark about the way participants process information that is inconsistent or else consistent with presupposed content. Participants significantly slowed down (by 135 ms) when the fifth segment of the vignette is inconsistent with a prior information carried by the presupposition as opposed to when it is consistent with it. Second, and similarly to the Shetreet et al. cases, for those items in which the second sentence does not provide supporting presuppositional content for the idiom in the third sentence (e.g., when students in the course in item [10] all know each other before *breaking the ice* is used), we also find significant slowdowns compared to cases where the content in the vignette is consistent with the presupposition of the idiom, even though the spread is smaller (slowdowns are about 74 ms). Third, the fifth segments following nonsense idioms prompted by far the slowest reactions (135 ms slower than fifth-segments of vignettes that included conventional idioms but without presuppositional support and 210 ms slower than fifth-segments that included conventional idioms with presuppositional support).

These data are consistent with other recent work (Beck and Weber, 2020) that shows how an idiom, when used in a context that biases a participant toward a "high literality" reading, prompts slowdowns two segments after the idiom. For example, the segment "sooner than later" in (13b) prompts slowdowns compared to cases in which the same segment appears after a figurative meaning is encouraged (13a).

> (13) (a) The new schoolboy/ who did not know/anyone in his class/just wanted to/break the ice/with his peers/sooner than later/ Monday morning.

> (b) The chilly Eskimo/ who was eager/to catch some fish/just wanted to/break the ice/with his peers/sooner than later/ Monday morning.

Our presuppositional account would argue that the presuppositional content of *break the ice* is unsatisfied in (13b) but satisfied in (13a).

We note that data from Griffen and Noveck (2023) aim to address cases in which participants receive brief stories. It is not clear how our account can be extended to cases where participants need to make grammatical judgments of idioms versus yoked controls. In our view, the speed advantages for idioms on grammatical judgment tasks might well depend on other features of idioms such as frequency and familiarity (e.g., see Libben and Titone, 2008; Carrol and Conklin, 2020).

## 6. Conclusion

This paper began by describing how idiom processing appears to be exceptional in the context of figurative language literature. The *idiom superiority effect* intriguingly reports faster reading times for idiomatic readings when compared to their literal controls, providing this literature with a unique effect when compared to other investigated figures, such as ironical and metaphorical readings (when compared to their controls). Essentially, we were driven to better understand this paradox. We thus sought to account for the characteristics that idioms tend to hold as we proposed a relevance-theoretic interpretation on their processing. For this reason, we reviewed a relevance-theoretic account offered by Vega-Moreno (2003, 2005), where idioms are treated as conventional metaphors. While we found Vega-Moreno's approach of treating idioms as conventional metaphors enriching, it did not provide the wherewithal to account for the *idiom superiority effect*.

We subsequently went on to propose that idiomatic strings generate a set of background assumptions, which verge on being presuppositional. This implies that each idiom is considered individually and that each idiomatic string necessitates specific contextual conditions in order to be considered felicitous. It follows that an idiomatic string will appear felicitous if there is contextual support and it will prompt incongruity if there is no contextual support. Our RT-inspired work leads to the conclusion that – if the idiom string is recognized as such and processed as a whole – it will automatically activate some procedure leading to the recovery of its associated background assumptions. Given that these assumptions conflict with the context when this is intended to support a literal interpretation, processing difficulties are to be expected. In other words, according to our approach, idiomatic interpretations will tend to be produced even if they are being used in literal control conditions. In this way, the *idiom superiority effect* is a natural consequence of our analysis. Of the three prominent idiom processing accounts in the psycholinguistic literature (as described in the Introduction) – the *compositional, direct retrieval* and *hybrid* accounts – our proposal is most compatible with the last two because we would have to assume that idiomatic meanings are generated (at some point) even in the literal control conditions, thus producing incongruities and slowdowns.

Turning to the RT literature, our explanation of idioms and their role in the *idiom superiority effect* is actually consistent with prior analyses of some key properties of procedural meaning. Specifically, Escandell-Vidal and Leonetti (2011) have suggested that procedural meaning is characterized by so-called 'rigidity', so that "procedural meaning will always prevail (i.e., impose its conditions) even when it enters into contradiction with other kinds of information, both linguistically encoded and contextually inferred" (2011, p. 81). Based on this, they maintain that for interpretation to succeed, the instructions encoded by an item must be satisfied and, as a consequence, any possible mismatch must be resolved by preserving the representations obtained by following the instructions. This rigidity is evident in the many cases of idiomatic expression that we discussed and that lead to mismatching contexts. In the example *Has the love of my life, my inspiration, Tom, kicked the bucket yet*? (which is a modification of [6]), it is interesting to notice that, as the range of assumptions and attitudinal disposition recovered by following the procedure encoded by the idiom cannot be overridden by the conceptual information encoded by the expressions *love of my life* and *my inspiration*; the only possible way to achieve a sensible and relevant interpretation of the utterance is to adopt an ironical interpretation of these expressions. Such an ironical interpretation would preserve the background assumptions derived by the application of a rigid procedure and resolve the mismatch at issue.

In sum, through the examples that we have provided and the preliminary results from both linguistic-intuition-based tests and ongoing experimental work, we hope that this paper can serve as an introductory, albeit convincing, argument for viewing idioms as a class of unique figurative expressions with their own processing requirements. One of these, which has been largely overlooked, is the way idioms include a procedural meaning that takes into account relevant presuppositional information. Once this feature of idioms is taken into consideration, one is in a better position to account for the *idiom superiority effect*. As we have outlined in this paper, our next step will be to follow up on our pilot experiment. It is our hope that our novel approach to idiom comprehension will enrich discussions of figurative language and its processing.

## Data availability statement

Further inquiries about original findings can be directed to the second author or the corresponding author.

## Author contributions

The authors made the following contributions. IN: conceptualization, writing – original draft preparation, writing – review and editing, experiment conceptualization and preparation, project administration, supervision, and methodology. NG: conceptualization, writing – draft preparation, editing, experimental

preparation, and data analysis. DM: conceptualization, writing – original draft preparation, writing – review and editing, and relevance theoretic analyses. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by seed money from the Laboratoire d'excellence Empirical Foundations of Linguistics (Labex-EFL) in Paris as well as from a Chaire d'Excellence grant from the Université de Paris-Cité to IN, a doctoral grant from the Université de Paris to NG, and an SNSF Eccellenza Grant (PCEGP1_186931/1) to DM.

## 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.

## Publisher's note



## References

Almor, A., Arunachalam, S., and Strickland, B. (2007). When the creampuff beat the boxer: working memory, cost, and function in reading metaphoric reference. *Metaphor. Symb.* 22, 169–193. doi: 10.1080/10926480701235478

Beaver, D. I., Geurts, B., and Denlinger, K. (2021). "Presupposition" in *The stanford encyclopedia of philosophy* (*Spring 2021 Edition*). ed. E. N. Zalta.

Beck, S. D., and Weber, A. (2020). Context and literality in idiom processing: evidence from self-paced reading. *J. Psycholinguist. Res.* 49, 837–863. doi: 10.1007/s10936-020-09719-2

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D. (2002). *Relevance and linguistic meaning: the semantics and pragmatics of discourse markers* Cambridge: Cambridge University Press.

Blakemore, D. (2011). On the descriptive ineffability of expressive meaning. *J. Pragmat.* 43, 3537–3550. doi: 10.1016/j.pragma.2011.08.003

Bobrow, S. A., and Bell, S. M. (1973). On catching on to idiomatic expressions. *Mem. Cogn.* 1, 343–346. doi: 10.3758/BF03198118

Cacciari, C., and Tabossi, P. (1988). The comprehension of idioms. *J. Mem. Lang.* 27, 668–683. doi: 10.1016/0749-596X(88)90014-9

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., and Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: an EEG study. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 928–943. doi: 10.1037/xlm0000351

Carrol, G., and Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Lang. Speech* 63, 95–122. doi: 10.1177/0023830918823230

Carrol, G., and Littlemore, J. (2020). Resolving figurative expressions during reading: the role of familiarity, transparency, and context. *Discourse Process.* 57, 609–626. doi: 10.1080/0163853X.2020.1729041

Carston, R. (2002). Linguistic meaning, communicated meaning and cognitive pragmatics. *Mind Lang.* 17, 127–148. doi: 10.1111/1468-0017.00192

Carston, R. (2016). The heterogeneity of procedural meaning. *Lingua*, 175, 154–166.

Chierchia, G., and McConnell-Ginet, S. (1990). *Meaning and grammar: An introduction to semantics*. Cambridge: MIT Press.

Cutting, J. C., and Bock, K. (1997). That's the way the cookie bounces: syntactic and semantic components of experimentally elicited idiom blendsß. *Mem. Cogn.* 25, 57–71. doi: 10.3758/BF03197285

de Saussure, L. (2013). Background relevance. *J. Pragmat.* 59, 178–189. doi: 10.1016/j.pragma.2013.08.009

Escandell-Vidal, V., and Leonetti, M. (2011). On the rigidity of procedural meaning. Procedural meaning: Problems and perspectives, Victoria Escandell-Vidal, Manuel Leonetti and Aoife Ahern Bingley: Emerald Group Publishing Limited.

Espinal, M. T., and Mateu, J. (2019). "Idioms and phraseology" in *Oxford research encyclopedia of linguistics*. eds. M. T. Espinal and J. Mateu (Oxford: Oxford University Press)

Ferretti, T. R., Singer, M., and Harwood, J. (2013). Processes of discourse integration: evidence from event-related brain potentials. *Discourse Process.* 50, 165–186. doi: 10.1080/0163853X.2013.766123

Ferretti, T. R., Singer, M., and Patterson, C. (2008). Electrophysiological evidence for the time-course of verifying text ideas. *Cognition* 108, 881–888. doi: 10.1016/j.cognition.2008.06.002

Fraser, B. (1970). Idioms within a transformational grammar. *Found. Langua.* 6, 22–42.

Gibbs, R. W. (1986). Skating on thin ice: literal meaning and understanding idioms in conversation. *Discourse Process.* 9, 17–30. doi: 10.1080/01638538609544629

Gibbs, R. W. (1991). Why idioms mean what they do. *J. Exp. Psychol. Gen.* 120, 93–95. doi: 10.1037/0096-3445.120.1.93

Gibbs, R. W., and Colston, H. L. (2012). *Interpreting figurative meaning*. Cambridge: Cambridge University Press.

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Mem. Cogn.* 8, 149–156. doi: 10.3758/BF03213418

Gibbs, R. W., Nayak, N. P., and Cutting, C. (1989). How to kick the bucket and not decompose: analyzability and idiom processing. *J. Mem. Lang.* 28, 576–593. doi: 10.1016/0749-596X(89)90014-4

Giora, R. (1997). Understanding figurative and literal language: the graded salience hypothesis. *Cogn. Linguis.* 8, 183–206. doi: 10.1515/cogl.1997.8.3.183

Giora, R., and Fein, O. (1999). Irony: context and salience. *Metaphor. Symb.* 14, 241–257. doi: 10.1207/S15327868MS1404_1

Giora, R., Fein, O., Laadan, D., Wolfson, J., Zeituny, M., Kidron, R., et al. (2007). Expecting irony: context versus salience-based effects. *Metaphor. Symb.* 22, 119–146. doi: 10.1080/10926480701235346

Giora, R., Fein, O., and Schwartz, T. (1998). Irony: grade salience and indirect negation. *Metaphor. Symb.* 13, 83–101. doi: 10.1207/s15327868ms1302_1

Glanzberg, M. (2003). Felicity and presupposition triggers. In University of Michigan Workshop in philosophy and linguistics.

Glucksberg, S., Gildea, P., and Bookin, H. B. (1982). On understanding nonliteral speech: can people ignore metaphors? *J. Verbal Learn. Verbal Behav.* 21, 85–98. doi: 10.1016/S0022-5371(82)90467-4

Griffen, N., and Noveck, I. (2023). Taking stock of an idiom's background assumptions: an alternative relevance theoretic account. [poster presentation]. Tenth biennial meeting of experimental pragmatics, Paris, France.

Ivanko, S. L., and Pexman, P. M. (2003). Context incongruity and irony processing. *Discourse Process.* 35, 241–279. doi: 10.1207/S15326950DP3503_2

Jorgensen, J., Miller, G. A., and Sperber, D. (1984). Test of the mention theory of irony. *J. Exp. Psychol. Gen.* 113, 112–120. doi: 10.1037/0096-3445.113.1.112

Jouravlev, O., Stearns, L., Bergen, L., Eddy, M., Gibson, E., and Fedorenko, E. (2016). Processing temporal presuppositions: An event-related potential study. *Langua. Cognition Neurosci.* 31, 1245–1256.

Karttunen, L. (1974). Presupposition and linguistic context. *Theor. Linguis.* 1, 181–194. doi: 10.1515/thli.1974.1.1-3.181

Katz, A. N., Blasko, D. G., and Kazmerski, V. A. (2004). Saying what you don't mean: social influences on sarcastic language processing. *Curr. Dir. Psychol. Sci.* 13, 186–189. doi: 10.1111/j.0963-7214.2004.00304.x

Langendoen, D., and Savin, H. (1971). "The projection problem for presuppositions" in *Studies in linguistic semantics*. eds. C. Filmore and D. Langendoen (New York: Holt)

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levorato, M. C., Nesi, B., and Cacciari, C. (2004). Reading comprehension and understanding idiomatic expressions: a developmental study. *Brain Lang.* 91, 303–314. doi: 10.1016/j.bandl.2004.04.002

Lewis, D. (1979). "Scorekeeping in a language game" in *Semantics from different points of view* (Berlin, Heidelberg), 172–187. doi: 10.1007/978-3-642-67458-7_12

Libben, M. R., and Titone, D. A. (2008). The multidetermined nature of idiom processing. *Mem. Cogn.* 36, 1103–1121. doi: 10.3758/MC.36.6.1103

Mancuso, A., Elia, A., Laudanna, A., and Vietri, S. (2020). The role of syntactic variability and literal interpretation plausibility in idiom comprehension. *J. Psycholinguist. Res.* 49, 99–124. doi: 10.1007/s10936-019-09673-8

Mazzarella, D., and Domaneschi, F. (2018). Presuppositional effects and ostensive-inferential communication. *J. Pragmat.* 138, 17–29. doi: 10.1016/j.pragma.2018.09.012

Noveck, I. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge: Cambridge University Press.

Noveck, I., Bianco, M., and Castry, A. (2000). The costs and benefits of metaphor. *Metaphor. Symb.* 16, 79–91.

Noveck, I., and Sperber, D. (2007). "The why and how of experimental pragmatics: the case of scalar inferences" in *Advances in Pragmatics Basingstoke*. ed. N. Roberts (London: Palgrave)

Nunberg, G. (1978). *The pragmatics of reference*. Bloomington:Indiana University Linguistics Club.

Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms, Idioms. *Language* 70, 491–538. doi: 10.1353/lan.1994.0007

Ortony, A., Schallert, D. L., Reynolds, R. E., and Antos, S. J. (1978). Interpreting metaphors and idioms: some effects of context on comprehension. *J. Verbal Learn. Verbal Behav.* 17, 465–477. doi: 10.1016/S0022-5371(78)90283-9

Pexman, P. M. (2008). It's fascinating research: the cognition of verbal irony. *Curr. Direct. Psychol. Sci.* 17, 286–290. doi: 10.1111/j.1467-8721.2008.00591.x

Pexman, P. M., Ferretti, T. R., and Katz, A. N. (2000). Discourse factors that influence online reading of metaphor and irony. *Discourse Process.* 29, 201–222. doi: 10.1207/S15326950dp2903_2

Potts, C. (2015). Presupposition and implicature. *Handb. Contemp. Seman. Theory* 2, 168–202. doi: 10.1002/9781118882139.ch6

Pynte, J., Besson, M., Robichon, F. H., and Poli, J. (1996). The time-course of metaphor comprehension: an event-related potential study. *Brain Lang.* 55, 293–316. doi: 10.1006/brln.1996.0107

Reinecke, R. (2020). Presuppositions: an experimental investigation (Doctoral dissertation, Université de Lyon).

Ronderos, C. R., Tomlinson, J., and Noveck, I. A. (2023). When irony is faster than its literal control: The role of mindreading during irony comprehension. *J. Exp. Psychol.*

Rubio-Fernandez, P. (2007). Suppression in metaphor interpretation: differences between meaning selection and meaning construction. *J. Semant.* 24, 345–371. doi: 10.1093/jos/ffm006

Schwarz, F. (Ed.). (2015). "Introduction: aspects of meaning in context - theoretical issues and experimental perspectives" in *Experimental perspectives on presuppositions* (New York, NY: Springer), 1–38.

Schwoebel, J., Dews, S., Winner, E., and Srinivas, K. (2000). Obligatory processing of the literal meaning of ironic utterances: further evidence. *Metaphor. Symb.* 15, 47–61. doi: 10.1080/10926488.2000.9678864

Shetreet, E., Alexander, E. J., Romoli, J., Chierchia, G., and Kuperberg, G. (2019). What we know about knowing: presuppositions generated by factive verbs influence downstream neural processing. *Cognition* 184, 96–106. doi: 10.1016/j.cognition.2018.11.012

Simons, M. (2005). "Presupposition and relevance" in *Semantics vs. pragmatics*. ed. Z. Szabó (Oxford: Clarendon Press)

Sperber, D., and Wilson, D. (1986/1995). *Relevance: communication and cognition*. Oxford: Blackwell.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., and Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: the case of irony. *Neuro Image* 63, 25–39. doi: 10.1016/j.neuroimage.2012.06.046

Spotorno, N., and Noveck, I. A. (2014). When is irony effortful? *J. Exp. Psychol. Gen.* 143, 1649–1665. doi: 10.1037/a0036630

Swinney, D. A., and Cutler, A. (1979). The access and processing of idiomatic expressions. *J. Verbal Learn. Verbal Behav.* 18, 523–534. doi: 10.1016/S0022-5371(79)90284-6

Tabossi, P., Fanari, R., and Wolf, K. (2009). Why are idioms recognized fast? *Mem. Cogn.* 37, 529–540. doi: 10.3758/MC.37.4.529

Titone, D., Lovseth, K., Kasparian, K., and Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement measures of reading. *Can. J. Exp. Psychol.* 73, 216–230. doi: 10.1037/cep0000175

Vega-Moreno, R. E. (2001). Representing and processing idioms. UCL Working Papers in Linguistics 13, 73–107.

Vega-Moreno, R. E. (2003). Relevance theory and the construction of idiom meaning. UCL Working Papers in Linguistics 15, 83–104.

Vega-Moreno, R. E. (2005). Idioms, transparency and pragmatic inference. UCL Working Papers in Linguistics 17, 389–425. 2005. 25

Wharton, T. (2003). Natural pragmatics and natural codes. *Mind Lang.* 18, 447–477. doi: 10.1111/1468-0017.00237

Wilson, D. (2003). Relevance and lexical pragmatics. *Ital. J. Linguist.* 15, 273–292.

Wilson, D. (2011). The conceptual-procedural distinction: past, present and future. *Proc. Mean. Prob. perspect.* 25, 3–31. doi: 10.1108/S1472-7870(2011)0000025005

Wilson, D., and Carston, R. (2007). *A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts* London: Palgrave.

Wilson, D., and Carston, R. (2019). Pragmatics and the challenge of 'non-propositional'effects. *J. Pragmat.* 145, 31–38. doi: 10.1016/j.pragma.2019.01.005

Wilson, D., and Wharton, T. (2006). Relevance and prosody. *J. Pragmat.* 38, 1559–1579. doi: 10.1016/j.pragma.2005.04.012

frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 29 September 2023
DOI 10.3389/fcomm.2023.1181805



OPEN ACCESS

EDITED BY
Kate Scott,
Kingston University, United Kingdom

REVIEWED BY
Alison Hall,
De Montfort University, United Kingdom
Caroline Jagoe,
Trinity College Dublin, Ireland

*CORRESPONDENCE
Pauline Madella
✉ pauline.madella@beds.ac.uk

RECEIVED 07 March 2023
ACCEPTED 30 August 2023
PUBLISHED 29 September 2023

CITATION
Madella P (2023) Relevance and multimodal prosody: implications for L2 teaching and learning. *Front. Commun.* 8:1181805. doi: 10.3389/fcomm.2023.1181805



# Relevance and multimodal prosody: implications for L2 teaching and learning

Pauline Madella*

School of Education and English, University of Bedfordshire, Bedford, United Kingdom

In this paper, I build on Scott's relevance-theoretic account of contrastive stress. Contrastive stress works as an extra cue to ostension in altering the salience of a particular constituent in an utterance and, as a result, the salience of one particular interpretation of that utterance. I draw on Scott's argument that contrastive stress does not encode procedural meaning. Contrastive stress is *un*predictable and, as such, it is in confounding the hearer's expectations that it draws his attention to the accented word and prompt his search for different interpretive effects. I argue that contrastive stress is interpreted purely inferentially precisely because it is one of many pointing devices. It is to be interpreted by virtue of its interaction with other paralinguistic behaviors, all of which being different aspects of the same ostensive act of communication. This leads me to focus on the gestural nature of contrastive stress working as an act of pointing, which, as an ostensive communicative behavior, conveys that *if you look over there, you'll know what I mean*. Finally, I present the implications of analyzing contrastive stress in its multimodal context—as prosodic pointing—for the teaching and learning of L2 prosodic pragmatics and the development of interpretive abilities in the L2 hearer's mind.

KEYWORDS

**relevance theory, contrastive stress, ostension, multimodal prosody, prosodic pointing, prosodic pragmatics, L2 pragmatics instruction**

## 1. Introduction

"It is not what you said, it's *how* you said it!" (Culpeper, 2011). In English, the prosodic contours of an utterance are central in the conveyance of speaker meaning. In this paper, I focus on one most conspicuous prosodic pattern: "contrastive stress" (Sperber and Wilson, 1986/1995; Scott, 2021). English makes extensive use of contrastive stress and co-speech visual information, which together "enhances linguistic input, distorts it, or replaces it, and sometimes even contradicts it" (Rost, 2016, p. 42). With its extra "oomph", contrastive stress draws the hearer's attention to one particular constituent of the utterance, often to result in contrastive reading. The syllable that carries the stress is signaled in upper case:

(1) SHE's always been the breadwinner.

While contrastive stress is ubiquitous in English, it is more or less accessible across languages (Ladd, 1996), but it remains a universal highlighting device: a vocal correlate of a pointing gesture (Sperber and Wilson, 1986/1995; Scott, 2021). This bears implications for L2 prosody and pragmatics development and pedagogy.

I begin in Sections 2 and 3 by building on Scott (2021) relevance-theoretic account of contrastive stress and further supporting her argument that contrastive stress is interpreted purely inferentially. In Section 4, I argue that this is largely due to contrastive stress being interpreted by virtue of its interaction with co-pointing behaviors and other "gestural accompaniments" (Jones, 1956), in its multimodal context. Contrastive stress is a special behavior because (1) it is the most conspicuous example

of multimodal prosody; one that English makes extensive use of, and (2) it is a vocal correlate of a pointing gesture. In Section 5.1, I focus on pointing as a "special" behavior, thereby bringing further evidence of why contrastive stress is special as a multimodal prosodic pattern par excellence. In Section 5.2, I demonstrate the pedagogical implications of my account of contrastive stress as prosodic pointing in the context of fine-tuning L2 hearer's relevance mechanisms and understanding the pragmatics of L2 prosody.

## 2. Contrastive stress in English

In spoken English, prosodic patterns can be intentionally used to convey pragmatic meaning or "speaker meaning" (Wilson and Wharton, 2006). One such prosodic phenomenon par excellence is so-called "contrastive stress" (Sperber and Wilson, 1986/1995; Scott, 2017a,b, 2021). In English, what is commonly referred to as contrastive stress, although the terminology may vary in the literature, e.g., "prosodic contrastive focus" (Dohen et al., 2007), "prosodic pointing" (Loevenbruck et al., 2009), "contrastive focus" (Wells, 2006), "contrastive accent" (Bolinger, 1961), "nuclear heavy stress" (Haugen, 1949), is the use of marked tonicity, as opposed to unmarked tonicity. English is an intonation language, or pitch accent language (Wells, 2006). This means that there is a general tendency in English for the main pitch accent or "nucleus" to fall on the stressed syllable of the final content word of an intonation phrase (IP), as in (2):

(2) I'd love a COffee.

(2) is a case of unmarked tonicity, or neutral nucleus placement, what Chomsky and Halle (1968) would refer to as "normal stress" and describe as predictable. If the final content word repeats information, the nuclear accent will be shifted to highlight the last new piece of information, as in (3b). If the last content word in the IP highlights new, contrastive information, as in (3c), it will be accented:

(3a) Would you like a COffee?
(3b) I'd LOVE a coffee!
(3c) I'd love a TEA!

In (3c), although the stress does fall on the last content word in the IP, i.e., tea, it is a case of marked tonicity, as it serves a contrastive function. For the nuclear accent to result in contrastive reading in (3c), some other unexpected element(s) would be added, such as a change in tempo, loudness, and duration. Stress is generally understood as "greater auditory prominence" (Katamba, 1989, p. 221–242); it is realized with "greater articulatory care" (Gussenhoven, 2004, p. 15). Contrastive stress is described as the most conspicuous accent of all (Bolinger, 1961, p. 83). Its extra oomph is produced by conveying "acoustic salience" through "increased intensity and duration" (Ladd, 1996, p. 58). "Loudness" is indeed presented as one of its distinctive traits (Bloomfield, 1933; Jones, 1956; Katamba, 1989; Wells, 2006). In marked tonicity cases, the nuclear accent can fall on "virtually any word which the speaker chooses to highlight" (Katamba, 1989, p. 242). This echoes Bolinger's (1961, p. 96) argument that "one cannot predict with precision when, where, and how the shift will occur", making the location of the nucleus highly unpredictable, unless we are mind-readers (Bolinger, 1972). Consider how movement of the nucleus placement in the below utterances (4b—8b) results in the speaker producing different realizations (Clark, 2013) of one same sentence:

4(a) Is this the play you have been looking for?
4(b) THAT is the play I have been looking for.

5(a) Is this the play you have been looking for?
5(b) That IS the play I have been looking for.

6(a) Is this the book you have been looking for?
6(b) That is the PLAY I have been looking for.

7(a) Is this the play Gem' has been looking for?
7(b) That is the play **I** have been looking for.
In (7b), the nuclear placement in signaled in bold.

8(a) Is this the play you have been looking for?
8(b) That is the play I WAS looking for.

In 4b—8b, the nuclear accent falls on an element of the utterance that is not typically accented yet, in so doing, reflects the speaker's intention to produce meaningful effects. The accenting of a contrasting element draws the hearer's attention and guides him in working out the speaker's intended meaning. For example, in (7b), emphatic stress falls on "I" as opposed to Gem' and thereby prompts the hearer to look for extra meaningful effects and infer that it is her, not Gem', that had been looking for the play. The above examples (4b-8b) show that contrastive stress is used to draw attention to a constituent that is made to stand out for the hearer to believe that it bears some relevance to him and is worth processing.

In so-called marked tonicity, stress *per se* does not bear contrastive meaning (Scott, 2021). Scott's account resonates with Bolinger's (1961, p. 84) point that contrast is not a property of the accent itself but rather one of its functions being to "MEAN contrast". By using contrastive stress, the speaker only guides, re-focuses, or re-directs the hearer's attention, which results in a contrast. Dohen et al. (2007, p. 221) note that what they call "prosodic contrastive focus" is used to "emphasize a word or group of words in an utterance as opposed to another". Thus, contrastive stress necessarily results in a contrast between the focused object and what has been deliberately left unaccented or deaccented. Not only does Scott (2021, p. 39) agree in arguing that contrastive stress does not encode contrastive meaning, but she goes further and argues that, in fact, it does not encode anything. In other words, the interpretation of contrastive stress is done purely inferentially. In so doing, Scott's relevance-theoretic account of how contrastive stress is interpreted offers further insights into the nature of the inferential processes at play when processing and interpreting contrastive stress.

## 3. The relevance of contrastive stress

Sperber and Wilson's relevance theory goes along with Grice's (1967, p. 37) idea that "the very act of communicating creates expectations which it then exploits". As such, an act of communication conveys to the hearer that paying attention to it will be worth their while. This is the basis for the Communicative

Principle of Relevance (Sperber and Wilson, 1986/1995, p. 260), defined in (9).

(9) Communicative Principle of Relevance: Every act of ostensive communication communicates a presumption of its own optimal relevance.

Knowing how a hearer is likely to respond, the speaker can easily manipulate the effort to which the hearer is put and manipulate his expectations so as to trigger his search for effects which justify that effort (Sperber and Wilson, 1986/1995; Scott, 2017a,b, 2021). Prosodic patterns are used by the speaker so as to trigger the hearer's search for relevance and his expectation of positive cognitive effects:

> A communicator who wants some prosodic feature of her utterance to be understood as contributing to her meaning should therefore do her best to make it salient enough, and rich enough in effects, to be picked out by the relevance-theoretic comprehension procedure and help make the utterance relevant in the expected way (Wilson and Wharton, 2006, p. 442).

As an extra cue to ostension (Scott, 2017a,b, 2021), an unexpected prosodic feature, such as contrastive stress, comes with the presumption that it is salient enough and rich enough in effects to be worth attending to and processing. Contrastive stress is salient not just because of the nuclear accent itself but due to the unexpectedness of the prosodic placement:

> Any departure from neutral (or "expected") prosody would increase the hearer's phonological processing effort but would thereby encourage him to look for extra (or different) effects (Wilson and Wharton, 2006, p. 448).

In operating as an extra cue to ostension, contrastive stress comes with the presumption of its optimal relevance. It draws the hearer's attention to what he would have otherwise ignored, and focuses it on her intentions. This entails that the hearer is put to more effort only to raise his expectations of more or different cognitive effects (Scott, 2017a,b, 2021). Thus, contrastive stress does not lead to a "quick and cheap" inference (Tomlinson and Bott, 2013, p. 3569). It primarily *re*-focuses the hearer's attention, which causes it to be effort-*ful*. As House (2006, p. 1547) notes, "assigning salience orients the hearer to update her cognitive environment in a particular way". The updating of his cognitive environment or re-focusing of his attention necessarily involves extra processing effort on his part, which, concomitantly, raises the addressee's expectations of extra or different effects on the account that the speaker must have good reasons for re-orienting him in a particular way. Wilson and Carston (2019, p. 4) address precisely this point:

> In language use, departures from expected syntax, wording or prosody [. . . ] provide possible cues to ostension, focussing attention on particular aspects of the ostensive act and encouraging a search for additional interpretive effects.

As the most conspicuous accent of all, contrastive stress naturally stands out. It results in contrastive reading; however, it does not *encode* contrast. As Scott (2021, p. 39) argues, contrastive stress is purely inferentially interpreted. As she explains, it is the disconfirmation of the addressee's expectations that triggers his search for different cognitive effects. In "confounding" the addressee's expectations, contrastive stress invites the hearer to follow a contrastive inferential route. Another point that Scott (2021) puts forward to support her argument is that contrastive stress does not activate the same procedure each and every time it is used, and so it cannot be said to encode procedural meaning. As Wilson (2016, p. 17) notes, it "merely point[s] the addressee in the right direction rather than providing a full concept as a starting point for inference". In other words, contrastive stress does not provide conceptually encoded content in the way that content words do, nor does it provide the addressee with a specific and systematic procedural instruction for him to follow (Fretheim, 2002) in the way that reference assignment does. Unexpected prosodic placement can be said, however, to have an impact on what Sax (2011, p. 378) names "procedures of comprehension", but it does not encode procedural constraints. Contrastive stress is unpredictable (Bolinger, 1972) in that it is a reflection of the speaker's choices as to what part of the utterance should be rendered more salient on the basis of the meaning she is intending to convey on that occasion and the inferential route the hearer needs to follow to arrive at the speaker's intended interpretation.

## 4. From contrastive stress to prosodic pointing

In this paper, I build on Scott (2021) relevance-theoretic account of contrastive stress. I draw on her argument that contrastive stress does not bear contrastive meaning nor encode procedural meaning. I support her account by suggesting that contrastive stress cannot be said to encode this or that procedural instruction precisely because it is interpretable by virtue of its interaction with co-pointing devices and other "gestural accompaniments" (Jones, 1956), provided that these are available to the hearer. In face-to-face communication, utterances generally are composites of a range of different behaviors, all of which being integral parts of the ostensive act of communication. As Ladd (1996, p. 40) notes, it is difficult to "unravel prosody from its paralinguistic context". As Wharton (2016, p. 5) also points out:

> The parallels are so strong that a single, homogeneous account of these para-/non-linguistic behaviors seems to be required, one that embraces the fact that they are, for the most part, closely interlinked.

Psychologist McNeill (1985, p. 350) also describes those concomitant paralinguistic elements as "parts of a single psychological structure". It follows that contrastive stress must be considered in its multimodal context. Contrastive stress is not just a prosodic phenomenon; it is a multimodal phenomenon par excellence. Along with its gestural counterparts, contrastive stress plays an active part in "catching someone's eye, touching them, pointing, showing them something" (Wilson and Carston, 2019, p. 34). Contrastive stress is special precisely because it is probably the best illustration of *multimodal prosody*. Intonation in general, and contrastive stress specifically, is typically produced and interpreted

together with visual cues that play a crucial part in how prosodic patterns—in particular, unexpected prosodic patterns—are to be interpreted, as (10) illustrates:

(10) I did not know SHE was coming.

In (10), the words themselves come short of conveying the speaker's full intended meaning. The accenting of "she" is only one aspect of a larger gestural act of communication. To reach a hypothesis about the speaker's meaning, the addressee will likely process and incorporate the speaker's eye-, chin-, and head-pointing toward "she", her frown and a face and tone of voice that show disapproval or discontentment, all contributing to revealing the attitude of the speaker and how the words are to be interpreted. As Stevick (1982, p. 163) expresses: "Nonverbal communication provides the surface on which the words are written and against which they must be interpreted". In (10), the speaker communicates much more than what is said: The way it is said conveys that she is not particularly pleased to see that "she" is there, she's not friendly with "her", etc. The co-pointing modalities coincide harmoniously with the vocally-conveyed highlighting of "she", which is in line with research that has shown how "nods, hand gestures, and eye contact coincide very precisely with events in the spoken message" (Kendon, 1972; Ladd, 1996, p. 34). Beyond the decoding of the linguistic form, the para-linguistic features are salient enough to be read as relevant inputs to inferential processing, and it is on the basis of how they interact that the addressee is able to construct a hypothesis about the speaker's intended meaning by incorporating the pieces of the puzzle. These pieces or individual modalities may well be conceptual, but they will need to be adjusted in the process of interpreting the utterance through inferential work (Sperber and Wilson, 2015).

The very nature of utterances is complex, and what they communicate can be best described as "nebulous, contextually shaded and hard to pin down in conceptual terms" (Wharton, 2009, p. 146). As Madella and Wharton (2023) argue, it is by virtue of their interaction that the encoded concepts carried by individual modalities, such as a frown for disapproval, a nod for agreement, the vocal highlighting of a pronoun, eye-pointing are expected to be "adjusted or modulated in the course of the interpretation process" for the purpose of making one particular inference on that one particular occasion (Sperber and Wilson, 2015, p. 145). Scott (2021) makes a point that these modalities do not trigger the same procedural constraints each and every time they are used. It is indeed on the basis of its interaction with co-pointing modalities that its meaning is constructed, and so it is worked out purely inferentially. Sperber and Wilson's (1986/1995) theory of utterance interpretation involves going beyond the Gricean notion of speaker's meaning to accommodate the interpretation of vague and weaker communication. This is summarized by Wilson and Carston (2019, p. 34):

> Relevance theorists set out from the start to look for a set of pragmatic principles and mechanisms that can deal with the full range of overtly intentional communicative acts: verbal and non-verbal, showing and telling, determinate and indeterminate, literal and figurative, propositional and non-propositional.

While non-verbal ostensive behaviors can be used to infer determinate, strong interpretations, they are often associated with vague, non-propositional, thus weaker communication, where it is difficult for the hearer to pinpoint one definite inference. In such cases, the speaker does not commit to one single interpretation but rather make "an array of roughly similar conclusions" available to the hearer (Wilson and Wharton, 2006, p. 1569). As contrastive stress is used as part of a wider range of composites all participating in the act of "showing", pointing the hearer in the intended direction, its interpretation is bound to often be more of a "diffuse impression" (Wilson and Wharton, 2006, p. 1569). This is also more generally conveyed by cognitive scientists and psychologists Tomasello et al. (2007, p. 705) when they write that:

> Pointing (…) does not convey a specific meaning in the manner of most conventionalized, symbolic gestures. Rather, pointing can convey an almost infinite variety of meanings by saying, in effect "If you look over there, you'll know what I mean".

Considering contrastive stress in its multimodal context and, therefore, as purely inferentially interpreted, assumes a natural account of prosody. The pragmatic nature of prosody comes through from the intimate connections it entertains with gesture (Bolinger, 1983a,b,c). In other words, it is in its gestural dimension that the pragmatics of prosody shows; it is precisely where its pragmatic force lies and what makes its pragmatic nature visible. Gesture is what brings prosody and pragmatics together; it bridges the gap between prosody and pragmatics by reflecting the gestural dimension and pragmatic force of prosody (Madella, 2021). I follow a natural approach to prosody (Bolinger, 1983a,b,c), thereby presenting contrastive stress as a natural highlighting device and illustrating Bolinger's point that speech prosody is one part of a broader "gestural complex". I focus on what I call *prosodic pointing*, or contrastive stress as one audio-visual construct. Adopting a natural approach to prosody, I contend that it is read the same way as gesture (Bolinger, 1983a,b,c) and treat prosody *as* gesture (Madella, 2021; Madella and Wharton, 2023). Thus, my perspective assumes a natural or universal approach to prosody, one that is in line with Bolinger's (1964) view of intonation as existing "around the edge of language". The nature of prosody has been described as ranging from "natural" to purely linguistic (Wharton, 2009). Prosody has a dual nature (House, 2006), so prosodic meaning is best described as a matter of degree rather than an all-or-nothing distinction reflected in either a natural or grammatical account. Bolinger strongly favors the idea that although we may feel some aspects of intonation to be linguistic, those aspects retain a degree of naturalness and can easily be traced back to their natural origins:

> Intonation… assists grammar—in some instances may be indispensable to it—but it is not ultimately grammatical… If here and there it has entered the realm of the arbitrary, it has taken the precaution of blazing a trail back to where it came from.

I, too, as far as contrastive stress is concerned, favor the view of prosody as a largely natural phenomenon, which belongs in the realm of pragmatics. This view contributes to our

understanding of contrastive stress as a multimodal phenomenon interpreted inferentially.

As a conclusion to his cross-linguistic study of accentuation variation, Ladd (1996, p. 167) argues against the idea of "some universal highlighting function" of prosody, which I disprove for reasons which will become apparent. The picture is indeed more complex as variability of accentuation is not consistent across languages. It is language specific (Sperber and Wilson, 1986/1995; Scott, 2021) and conditioned by the grammatical constraints of specific languages. In Norwegian, for instance, Fretheim (1998) explains that the word-accent system severely restricts the communicator's intonational patterns. As illustrated in Section 2, English allows for flexible prosodic placement so long as it contributes to the speaker conveying her intended meaning. In other words, pragmatics can prevail over strict structural considerations. English is said to enjoy high pragmatically-motivated accentuation variability (Madella, 2021). This also suggests that pitch-marked prominence in English is more subject to unpredictability and a reflection of the speaker's choices in comparison with languages that rely more heavily on structural constraints in their accent placement. Ladd's (1996) shows that contrastive stress is less accessible cross-linguistically, while it is ubiquitous in English. However, this should not lead to the conclusion that the study fails to reveal the universal nature of contrastive stress. From a relevance-theoretic perspective, it shows that contrastive stress is more or less disruptive across languages (Wilson and Wharton, 2006; Wharton, 2009; Scott, 2017a). It will be less accessible to speakers of languages which do not place focal stress as freely as English does and make use of other, syntactic, constructions. French, for example, more typically uses cleft forms, as in (11a–c) below. The asterisk indicates that the utterance is ungrammatical or not typically used:

(11a) C'est elle qui l'a fait. *It is her who did it.
(11b) C'est ELLE qui l'a fait. *It is HER who did it.
(11c) *ELLE l'a fait. SHE did it.

The syntactic extraction illustrated in (11a) is preferred over stressing "elle" to mark focus in French. This is not to say that contrastive stress in French is not at all possible, but cleft constructions are generally preferred. In (11b), both syntactic and prosodic contrastive focus (Dohen et al., 2007) are used. It is, however, used more sparsely and the accent is not quite equivalent to the intensity, duration, and loudness that characterize contrastive stress in English. That is due to the cleft construction contributing more heavily to the highlighting of the pronoun. French prosodic patterns do not allow for contrastive stress to be used as easily as it is used in English (Scott, 2021), and French has other preferred ways of conveying pointing, such as syntactic pointing. This is partly due to French being a non-intonation language. Similarly, in Spanish, the "a él" structure in (12a) would be preferred over the accenting of "lo" in (12b) (VanPatten, 2018):

(12a) Bill lo conoce a él.
(12b) *Bill LO conoce.
(12c) Bill knows HIM.

The syntactic construction "a él" in (12a) will more likely be used to highlight "lo" (i.e., "him"). "A él" is the Spanish syntactic equivalent of prosodic stress on "him" in English (VanPatten, 2018). Another example of accentuation variability across languages is Italian (Ladd, 1996). Italian is known as a +rightmost language along with other languages, such as Spanish and Romanian. These languages resist *deaccenting*. In English, the accenting of an element that is typically unmarked necessarily entails that an element which would have been expected to carry the accent consequently becomes deaccented. The extensive use of contrastive stress contributes to deaccenting being an ordinary pattern in English, for example, in cases of repeated or given information. Semantic weight, semantic impoverishment, and semantic emptiness all are further conditions for deaccenting in English. However, +rightmost languages, like Italian, generally resist deaccenting of repeated material, empty content words, or last words:

(13) I made a TRIfle, but he HAtes desserts.

In the second intonation phrase of (13), "hates" rather than "desserts" would be accented, for "hates" is new information and so considered semantically richer as opposed to "desserts", which is information already given by "Trifle". In Italian, "desserts", i.e., the +rightmost word, would typically be accented. It does not follow from Ladd's study that contrastive stress cannot be regarded as a "natural" highlighting device across languages. What it does show is that contrastive stress is likely to be more or less disruptive across languages and, therefore, costlier and used more sparingly in those languages that rely more heavily on structural constraints (Sperber and Wilson, 1986/1995; Wilson and Wharton, 2006). As relevance-theorists argue, contrastive stress can be analyzed in terms of processing effort and cognitive effects. The process by which unexpected prosodic patterns put the hearer to extra processing effort and thus lead him to expect richer effects is universal (Wilson and Wharton, 2006; Scott, 2017a,b, 2021). The hearer is well aware that extra interpretive effects will likely offset the extra effort put in processing contrastive stress. In fact, contrastive stress is so routinely and ubiquitously used in English that it is expected to bear extra or different meaningful effects.

While Ladd (1996) demonstrates that the idea of intonation universals falls short in some way, the use of contrastive stress is often coupled with production of more universally recognized action, as demonstrated above. When Ladd (1996, p. 167) concludes that sentence accentuation is not "simply a matter of applying some universal highlighting gesture to individually informative words", he is not far from claiming that a showing gesture or gestural highlighting would likely be more universal and would thus be less controversially recognized as natural. Bolinger's description of a possibly pre-linguistic (almost biological) highlighting function of intonational contours used for the reading of speakers' mental states and intentions has been controversial. His description, however, seems to suit an arguably less controversial pre-linguistic (and certainly biological) universal of human communication: *pointing*. According to Scott (2021, p. 37), contrastive stress, as an ostensive behavior, operates much like a pointing gesture. The speaker is "pointing to a

part of the utterance with her voice" (Scott, 2021). Imai (1998) describes prosody as a relevance indicator, some sort of natural "pointer" indicating where relevance is to be found. As Sperber and Wilson (1986/1995, p. 203) put it, "stress is a sort of vocal equivalent of pointing [...] a natural means of drawing attention to one particular constituent in an utterance". Indeed, the deictic nature of contrastive stress makes it a very close equivalent to a pointing gesture. Scott's (2017a) characterization of contrastive stress or vocal pointing provides further elements of support to why contrastive stress should be treated as one of many pointing modalities. Contrastive stress and pointing are (extra) cues to ostension, raising expectations and producing non-encoded meaning (Scott, 2017a):

(1) They are both ostensive. They both prompts and guides the hearer's inferential work. They focus attention and focus it on the speaker's reasons for drawing his attention.
(2) They both raise expectations of the ostensive stimuli's optimal relevance. They both manipulate the hearer's expectations in confounding them and thereby triggering his search for additional or different cognitive effects, which will justify and offset the extra processing effort required to retrieve the intended interpretation.
(3) Both contrastive stress and a pointing gesture merely points the addressee in the intended direction without encoding anything. They are means of showing something and, in doing so, they guide the search for relevance.

Both contrastive stress and gestural pointing are driven by the same motivation (Madella, 2021); they are two aspects of the same process of utterance (Kendon, 1972, 1980). So based on Ladd's conclusive remarks, contrastive stress *can* be seen as a natural universal highlighting device, one that is typically used as one of and along with many other pointing devices (Wilson, 2016). The argument for considering contrastive stress as a multimodal prosodic phenomenon appears even stronger when we look at pointing and why it is a special behavior.

# 5. Prosodic pointing is special: implications for L2 prosodic pragmatics

## 5.1. Pointing is special

Pointing is indisputably "special", which makes contrastive stress a special multimodal phenomenon. For one thing, pointing lies at the root of human communication. It is ubiquitous and likely universal (Kita, 2003; Loevenbruck et al., 2008, 2009). A pointing gesture is typically performed "with the index finger and arm extended in the direction of the interesting object and with the other fingers curled inside the hand" (Butterworth, 2003, p. 9). Pointing in children is first expressed with both the eyes and the finger. It is then communicated via intonation, and finally with syntax. Ocular and manual forms of pointing are not the only way of expressing pointing through gesture, as example (10) has shown. Chin, eye gaze, and associated eyebrow motion could be added to the list, depending on which part of the world you are in. Lip-pointing, on the other hand, is not exactly common nor socially recognized around Europe, but it is a widespread deictic gesture in Southeast Asia, the Americas, Africa and Oceania. A study of Lao speakers' use of lip-pointing describes it as not only involving "protruding one or both lips, but also raising the head, sticking out the chin, lifting the eyebrows, among other things" (Enfield, 2001, pp. 185-191). In Māori gesture, eyebrow flashes are yet another specific form of pointing (Gruber et al., 2016).

Pointing is a "special" multimodal behavior in the brain as well. Loevenbruck et al.'s (2008; 2009) focuses on the more biological aspect of pointing and the cerebral domains that multimodal pointing recruits. They find that vocal and gestural pointing recruit similar cerebral domains; the two modalities are produced and perceived simultaneously (Loevenbruck et al., 2008, 2009). Loevenbruck et al.'s (2008; 2009) research on pointing is in line with the natural argument: If those pointing modalities entertain such intimate connections in the brain, it certainly shows that contrastive stress, as a paralinguistic, biological phenomenon should be discussed as one audio-visual construct. As they note, pointing, or a deictic behavior, is a "universal ability which orients the attention of another person so that an object/person/direction/event becomes the shared focus of attention" (Loevenbruck et al., 2008 p.1). The major role played by manual or indexical pointing in language development strongly suggests that "vocal pointing and pointing in other modalities may well be grounded in a common cerebral network" (Loevenbruck et al., 2008, p.1). This is also indicated by Hübscher and Prieto (2019), who describe gestural and prosodic development as "sister systems", operating in parallel in the brain and jointly contributing to L1 socio-pragmatic development. Dohen et al. (2007) and Loevenbruck et al. (2008) suggest that the detection and perception of contrastive stress—what they call prosodic contrastive focus—relies on the reading of multimodal cues. Dohen et al. (2007) reported the results of Tong et al.'s (2005) study of the neural processes underlying the perception of contrastive stress as opposed to that of intonation for question and affirmation discrimination. Their results indicated that processing contrastive stress involves more diffused neural activity. Dohen et al. (2007) compared French participants' perception of prosodic focus with that of syntactic pointing (used more typically in French). They found that processing syntactic pointing merely involved the frontal region of the brain, while processing prosodic contrastive focus—what I call contrastive stress—recruited *frontal and left parietal regions*. The left parietal regions are typically associated with other forms of pointing, such as gestural pointing. Perception and production of contrastive stress therefore seem to recruit multimodal activity. This was further supported by Dohen and Loevenbruck's (2009) study on the interaction of audition and vision for the perception of prosodic contrastive focus. Their study (Dohen and Loevenbruck, 2009, p. 7) demonstrated that:

> Even though the perception of prosodic focus is often considered as uniquely auditory, it is possible to perceive prosodic focus visually and the visual modality can enhance perception when prosodic auditory cues are degraded.

The above thus suggests that English speakers would recruit associative brain regions in their production and perception of contrastive stress. Dohen et al.'s work not only gives further motivation to look at contrastive stress as a gestural complex,

as one audio-visual construct, but it emphasizes the necessity to use multisensory information to detect contrastive stress in English and to consider the perception of contrastive stress as multimodal. The above neurological claims provide ample evidence that contrastive stress must be analyzed in its multimodal context, as a part of a broader audio-visual construct, which in turn offers further support to Scott's account according to which contrastive stress does not encode anything and is interpreted purely inferentially.

### 5.2. The relevance of prosodic pointing to L2 prosodic pragmatics development

As noted in Section 4, an analysis of contrastive stress as a multimodal prosodic phenomenon contributes to bridging the gap between prosody and pragmatics. Prosodic pointing—contrastive stress as one audio-visual construct—does a good job at illustrating the pragmatics of prosody or what Romero-Trillo (2012, 2016, 2019) calls *prosodic pragmatics*.[1] I have argued that the pragmatic force of prosody does not come from prosody alone. It lies in the gestural dimension of prosody and in the way that prosody naturally interacts with other paralinguistic communicative behaviors. I have argued and demonstrated (Madella and Romero-Trillo, 2019; Madella, 2021; Madella and Wharton, 2023) that analyzing contrastive stress as a multimodal construct bears important L2 pedagogical implications. Exposure to multimodal prosody generally, and prosodic pointing specifically, can be used toward fine-tuning L2 relevance mechanisms triggered by multimodal input to inferential processing. In other words, it enables L2 hearers to understand the speaker's non-verbal communicative behaviors as evidence of her intentions. It can therefore enhance L2 hearers'[2] ability and willingness to move beyond conceptual meaning and trust paralinguistic input in retrieving the speaker's intended interpretation. It was found that having access to prosodic pointing—after being exposed to contrastive stress alone—made L2 hearers appreciate the need to access multimodal input, for them to "remember more from visual information", "understand more clearly because (they) can see the body-language", and "see who speaks and their different faces" (Madella, 2021, p. 253, my amendment). Finally, it can develop the L2 hearer's alertness to the pragmatics of prosody and co-speech gesture and bodily accompaniments, which in turn contributes to the development of interpretive abilities in the L2 hearer. For instance, it was also found that the L2 hearer is more likely to understand the pragmatics of contrastive stress when it falls on "you" in the question "Would YOU like an apple?", if he also has access to the speaker's gestural behavior, i.e., leaning forward and using an open-palm hand gesture showing that she is returning a question.

1 Term used by Romero-Trillo (2012, 2016, 2019).

2 The term "hearer" remains as it follows from the relevance-theoretic tradition. It does not imply that the L2 hearer does not listen attentively, intentionally, or purposely.

## 6. Conclusion

In this paper, I have built on Scott (2021) relevance-theoretic account of contrastive stress and further supported her argument that, as an unpredictable extra cue to ostension disconfirming the hearer's expectations, contrastive stress is interpreted purely inferentially. I put forward the argument that it is precisely because contrastive stress is typically interpreted in its multimodal context that its meaningful effects are to be interpreted purely inferentially by virtue of its interaction with co-speech gesture and co-pointing modalities. As an ostensive behavior, contrastive stress operates the same way as a pointing gesture does, and the gestural nature of contrastive stress justifies that we want to analyse it in relevance-theoretic terms as *prosodic pointing*. Analyzing contrastive stress as a multimodal phenomenon—as prosodic pointing—further supports Scott's argument that contrastive stress does not encode procedural meaning. It simply points the hearer in the intended direction, where evidence of the speaker's intentions is to be found. Finally, analyzing contrastive stress as a multimodal phenomenon bears implications for the development and instruction of L2 prosody and relevance mechanisms (Madella and Romero-Trillo, 2019; Madella, 2021; Madella and Wharton, 2023).

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

## Acknowledgments

I am forever grateful to Tim Wharton for his continuing support. I would like to thank Frontiers and the editorial team for their trust and patience.

## 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.

The handling editor KS declared a past collaboration with the author PM.

## Publisher's note

# References


Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language* 37, 83–96. doi: 10.2307/411252

Bolinger, D. (1964). Around the edge of language: intonation. *Harv. Educ. Rev.* 34, 282–296. doi: 10.17763/haer.34.2.4474051q78442216

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language* 48, 633–644. doi: 10.2307/412039

Bolinger, D. (1983a). Where does intonation belong? *J. Semant.* 2, 101–120. doi: 10.1093/semant/2.2.101

Bolinger, D. (1983b). Intonation and gesture. *Am. Speech* 58, 156–174. doi: 10.2307/455326

Bolinger, D. (1983c). "The inherent iconism of intonation," in *Iconicity in Syntax*, ed J. Haiman. Amsterdam: John Benjamins, 97–109.

Butterworth, G. (2003). Pointing is the royal road to language for babies. *in Pointing: Where Language, Culture, and Cognition Meet*, ed S. Kita. Mahwah, NJ: Erlbaum, 9–33.

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clark, B. (2013). *Relevance Theory*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139034104

Culpeper, J. (2011). "It's not what you said, it's how you said it!": prosody and impoliteness," in *Discursive Approaches to Politeness. Mouton Series in Pragmatics [MSP]* (Berlin: Mouton de Gruyter), 57–83.

Dohen, M., and Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52, 177–206. doi: 10.1177/0023830909103166

Dohen, M., Loevenbruck, H., Callan, A., Callan, D., Baciu, M., Pichat, C., et al. (2007). *'Multimodal Perception of Prosodic Contrastive Focus in French: A Preliminary fMRI Study', Summerschool: Cognitive and physical models of speech production, perception and perception-production interaction*. Autrans, France: Part II*: Brain and Speech.

Enfield, N. J. (2001). 'Lip-pointing:' a discussion of form and function with reference to data from Laos. *Gesture* 1, 185–212. doi: 10.1075/gest.1.2.06enf

Fretheim, T. (1998). "Intonation and the procedural encoding of attributed thoughts: The case of Norwegian negative interrogatives," in *Current Issues in Relevance Theory*, eds V. Rouchota and A. Jucker. Amsterdam: John Benjamins, 205–236.

Fretheim, T. (2002). "Intonation as a constraint on inferential processing," in *Proceedings of the Speech Prosody 2002 Conference, Aix-en-Provence, France,* eds B. Bell and I. Marlien, 59–64.

Grice, P. H. (1967). "Logic and conversation" in *Unpublished MS from the William James Lectures*. Cambridge, Massachusetts: Harvard University.

Gruber, J., King, J., Hay, J., and Johnston, L. (2016). The hands, head, and brow: a sociolinguistic study of Māori gesture. *Gesture* 15, 1–36. doi: 10.1075/gest.15.1.01gru

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Haugen, E. (1949). Phoneme or prosodeme? *Language* 25, 278–282. doi: 10.2307/410089

House, J. (2006). Constructing a context with intonation. *J. Pragmat.* 38, 1542–1558. doi: 10.1016/j.pragma.2005.07.005

Hübscher, I., and Prieto, P. (2019). Gestural and prosodic development act as sister systems and jointly pave the way for children's sociopragmatic development. *Front. Psychol.* 10, 1259. doi: 10.3389/fpsyg.2019.01259

Imai, K. (1998). "Intonation and Relevance," in *Relevance Theory: Applications and Implications,* eds R. Carston and S. Uchida. Amsterdam: John Benjamins, 69–86.

Jones, D. (1956). *An outline of English Phonetics*. Cambridge, UK: CUP.

Katamba, F. (1989). *An Introduction to Phonology*. Harlow, Essex: Longman.

Kendon, A. (1972). "Some relationships between body motion and speech," in *Studies in Dyadic Communication*, eds A. Siegman and B. Pope. New York: Pergamon Press, 177–210.

Kendon, A. (1980). "Gesticulation and speech: two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, ed M. R. Key. The Hague: Mouton and Co., 207–227.

Kita, S. (2003) *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ: Erlbaum.

Ladd, R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Loevenbruck, H., Dohen, M., and Vilain, C. (2008). From gestural pointing to vocal pointing in the brain. *Revue Française de Linguistique Appliquée* 8, 23–33. doi: 10.3917/rfla.132.0023

Loevenbruck, H., Dohen, M., and Vilain, C. (2009). "Pointing is "special"", in *Some Aspects of Speech and the Brain,* eds S. Fuchs, H. Loevenbruck, D. Pape, and P. Perrier. Frankfurt am Main: Peter Lang, 211–258.

Madella, P. (2021). *Prosodic Pointing: from Pragmatic Awareness to Pragmatic Competence in Chinese Hearers of L2 English*. Brighton and Hove: University of Brighton.

Madella, P., and Romero-Trillo, J. (2019). Prosodic pointing in inferential comprehension: the application of relevance theory to L2 listening instruction. *Letrônica* 12, 1–17. doi: 10.15448/1984-4301.2019.4.34157

Madella, P., and Wharton, T. (2023). "Non-verbal communication and context: multi-modality in interaction," in *The Cambridge Handbook of Language and Context,* ed J. Romero-Trillo (Cambridge University Press), 419–435.

McNeill, D. (1985). So you think gestures are nonverbal? *Psych. Rev*. 92, 350–371. doi: 10.1037/0033-295X.92.3.350

Romero-Trillo, J. (2012) *Pragmatics, Prosody and English Language Teaching*. Dordrecht: Springer.

Romero-Trillo, J. (2016). "The pragmatics of prosody, the prosody of pragmatics [Plenary talk]," in *Séminaire de Recherche en Linguistique*. Geneva: University of Geneva, Switzerland.

Romero-Trillo, J. (2019). Prosodic pragmatics and feedback in intercultural communication. *J. Pragmat.* 151, 91–102. doi: 10.1016/j.pragma.2019.02.018

Rost, M. (2016). *Teaching and Researching Listening*. New York, NY: Routledge. doi: 10.4324/9781315732862

Sax, D. (2011). "Sentence Stress and the Procedures of Comprehension," in *Procedural Meaning: Problems and Perspectives (Current Research in the Semantics/Pragmatics Interface)*, eds V. Escandell-Vidal, M. Leonetti, and A. Ahern. Bingley: Emerald Group Publishing, 349–381.

Scott, K. (2017a). *Ostension, Expectations and Non-Encoded Meaning [Presentation]*. Belfast, UK: International Pragmatics Association (IPrA) Conference.

Scott, K. (2017b). "Prosody, procedures and pragmatics," in *Semantics and Pragmatics: Drawing a Line Series: Logic, Argumentation & Reasoning,* eds L. Depraetere and R. Salkie. Cham: Springer, 323–341.

Scott, K. (2021). "Contrastive stress in English: meaning, expectations and ostension," in *Pragmatics and Beyond New Ser. Beyond Meaning*, eds T. Wharton, L. D. Saussure, E. Ifantidou (Amsterdam: John Benjamins).

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

Sperber, D., and Wilson, D. (2015). Beyond speaker's meaning. *Croat. J. Philos.* XV, 117–149.

Stevick, E. W. (1982). *Teaching and Learning Languages*. Cambridge: Cambridge University Press.

Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x

Tomlinson, J. M., and Bott, L. (2013). "How intonation constrains pragmatic inferences," in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth, 3569–3575.

Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., et al. (2005). Neural circuity underlying sentence-level linguistic prosody. *NeuroImage* 28, 417–428. doi: 10.1016/j.neuroimage.2005.06.002

VanPatten, B. (2018). "The (risky) promise of instructed SLA [Presentation]," in *51st Annual Meeting of the British Association of Applied Linguistics (BAAL), York, UK.*

Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge: Cambridge University Press.

Wharton, T. J. (2009). *Pragmatics and Non-Verbal Communication*. Cambridge: Cambridge University Press.

Wharton, T. J. (2016). "Paralanguage," in *The Routledge Handbook of Pragmatics Routledge Handbooks in Applied Linguistics*, eds A. Barron, G. Steen, and G. Yueguo (London: Routledge), 69–75. doi: 10.1515/ip-2017-0032

Wilson, D. (2016). Reassessing the conceptual-procedural distinction. *Lingua* 175, 5–19. doi: 10.1016/j.lingua.2015.12.005

Wilson, D., and Carston, R. (2019). Pragmatics and the challenge of 'non-propositional' effects. *J. Pragmat.* 145, 31–38. doi: 10.1016/j.pragma.2019.01.005

Wilson, D., and Wharton, T. J. (2006). Relevance and prosody. *J. Pragmat.* 38, 1559–1579. doi: 10.1016/j.pragma.2005.04.012

frontiers | Frontiers in Psychology

TYPE Conceptual Analysis
PUBLISHED 29 January 2024
DOI 10.3389/fpsyg.2023.1167790



OPEN ACCESS

EDITED BY
Tim Wharton,
University of Brighton, United Kingdom

REVIEWED BY
Zhonggang Sang,
Xi'an Jiaotong University, China
Kate Scott,
Kingston University, United Kingdom

*CORRESPONDENCE
Marilynn Johnson
✉ marilynnjohnson@sandiego.edu

RECEIVED 16 February 2023
ACCEPTED 31 October 2023
PUBLISHED 29 January 2024

CITATION
Johnson M (2024) Relevance theory and the social realities of communication.
*Front. Psychol.* 14:1167790.
doi: 10.3389/fpsyg.2023.1167790



# Relevance theory and the social realities of communication

Marilynn Johnson*

Department of Philosophy, University of San Diego, San Diego, CA, United States

A central tenet of theories of meaning in the Gricean tradition—such as Relevance Theory—is that others will come to believe certain things simply by recognizing our intentions to communicate. In this article I demonstrate that this is not equally the case for all interlocutors; some bear additional burdens. In particular, I argue that this can happen in two ways. First, I demonstrate how a response to persistent testimonial injustice can be understood in terms of Sperber and Wilson's distinction between meaning-that and showing-that; a speaker who experiences repeated testimonial injustice will often move down the meaning vs. showing continuum. This is a result of a speaker learning that recognition of her intention has not in her experience been sufficient to induce the intended response in the hearer. Secondly, in consideration of social science research around perception of accent prestige and other status cues, I detail further costs borne by those who change their physical appearance and voice to be perceived as more credible. The costs of communication are not equal for all: they are greater for those who face a credibility deficit based in identity prejudice. Overall, by bringing Fricker's notion of testimonial injustice to bear on Relevance Theory, this article shows how social factors affect the reality of how interlocutors communicate.

KEYWORDS

**Relevance Theory, testimonial injustice, communication, Grice, Sperber and Wilson, Fricker, credibility**

## Introduction

A central tenet of theories of meaning in the Gricean, pragmatic tradition—such as Relevance Theory—is that others will come to believe certain things by recognizing our intentions to communicate. I argue that those working in this tradition need to consider the additional burden that is borne by some interlocutors in getting others to come to believe some content. I will demonstrate how a response to persistent testimonial injustice can be best understood in terms of a distinction presented by Dan Sperber and Deirdre Wilson between meaning-that and showing-that. I argue that a speaker who experiences repeated Testimonial Injustice may respond by moving down Sperber and Wilson's meaning vs. showing continuum. This explains an additional downstream effect not explicitly discussed by Miranda Fricker in her work on Testimonial Injustice. I then will present my understanding of what I call "social interpretation." In consideration of social science research around perception of prestige and status cues, I detail further costs borne by those who undertake the rational process of making changes to their physical appearance and voice to be perceived as more credible. I will conclude with what I see as the main takeaways from my argument.

This paper presents a socially-situated account of philosophy of language. In this focus, I follow work by philosopher Miranda Fricker who writes, "a socially situated account of a human practice is an account such that the participants are conceived not in abstraction from relations of social power…but as operating as social types who stand in relations of power to

one another" (Fricker, 2006, p. 3). This socially situated account stands in contrast to how much philosophy of language is usually conducted. It will come as no surprise to philosophers of language that the field has for the most part ignored these social realities of communication. But this might be news to those outside the debates within the discipline. For language is one of the ways that class, race, and power are most evident.

Ignoring this reality, examples in much philosophy of language literature are given in terms of "interlocutors" or discussions between people with names like "Smith" and "Jones," "Mary" and "Paul." What is the race of these interlocutors? What is their social status? What is their gender? Of course, in philosophy there is a certain amount of "compulsory rational idealization" that is necessary in presenting theoretical frameworks (Fricker, 2006, p. 2). However, in pragmatics—the branch of philosophy of language that seeks to turn away from abstract discussions of language itself and consider the reality of how we communicate with each other—we should aim to eventually turn away from abstraction and develop more fully-fleshed out accounts of the messy social realities that shape communication.

I will assume an intentionalist account of meaning as a starting place for this article. Of course, some reject an intentionalist account of meaning; but, defending intentionalism will not be my focus here (for such a defense see Johnson, 2019; Johnson, 2022a). Here my focus is a discussion of Relevance Theory, which falls within the Gricean, intentionalist tradition. I should also specify that my arguments here are not presented as a criticism of either the Sperber and Wilson or Fricker positions—but rather as a fruitful way of building on both of their theories by bringing them together.

## Meaning and showing

Let me now commence with presenting the relevant parts of Sperber and Wilson's theory. Sperber and Wilson's Relevance Theory was first presented in their 1986 *Relevance: Communication and Cognition* and they have continued to further develop their position since that time. One recent expansion on content from that book was their 2015 paper "Beyond Speaker's Meaning" in which they defend Relevance Theory broadly and develop further some theoretical machinery. Sperber and Wilson argue that their theory best captures what we want from a theory of communication—i.e. is more "conceptually unified," picks out "the proper object of a philosophical definition or a scientific theory," and "makes good sense of our fuzzy intuitions about speaker meaning" (Sperber and Wilson, 2015, p. 117).

Relevance Theory can be seen as a part of the Gricean tradition in that it follows in the footsteps of philosopher of language H. P. Grice, whom Deirdre Wilson studied with at Oxford. Relevance was one of Grice's proposed four maxims of conversation but the way relevance is understood by Sperber and Wilson is quite different. For them relevance is the key to ostensive-inferential communication. As they write, "By producing an utterance, the speaker requests the hearer's attention. By requesting his attention, she suggests that her utterance is relevant enough to be worth his attention. This applies not just to speech but to all forms of ostensive communication" (Sperber and Wilson, 1986, p. 154). By ostensive communication they mean other nonverbal acts such as pointing to a clock, or ringing a doorbell (Sperber and Wilson, 1986, p. 53; Sperber and Wilson, 2015).

Although they are part of the Gricean, pragmatic tradition, Sperber and Wilson depart from Grice in a number of other important ways (see Sperber and Wilson, 1986, p. 161–163, Carston, 2005, and Horn, 2006 for further discussion of how Relevance Theory relates to Grice). In presenting his account of meaning, Grice argues for a definition with three conditions—including a third clause that recognition of the speaker's intention be in some way *the basis* for a hearer to produce the intended response. In contrast, Sperber and Wilson prefer to work with a more "permissive" account that drops this requirement leaving only the first two. With the third clause dropped this picks out what they call "ostensive communication." They write,

> In characterising ostensive communication, we built on the first two clauses of Grice's definition and dropped the third. This was not because we were willing to broaden the definition of utterer's meaning—we agreed with Grice that talk of 'meaning' is awkward in certain cases—but because it seemed obvious that there is a continuum of cases between 'meaning that' (typically achieved by the use of language) and displaying evidence that (in other words, showing) and we wanted our account of communication to cover both (Sperber and Wilson, 2015, p. 119).

Whether it is better to work with Grice's original three clauses or drop them in favor of two is something that is debated amongst those working in the pragmatic tradition.[1]

By dropping Grice's third clause, Sperber and Wilson open up the sorts of relevant cases to include "meaning that" as well as "showing that," which they then define. They write that meaning that (MT) is "typically achieved by the use of language" and that showing that (ST) is "displaying evidence that" (Sperber and Wilson, 2015, p. 119). By dropping Grice's third clause the Sperber and Wilson account covers a wider range of communicative acts, including those cases where the intention to communicate is superseded by the direct evidence. For instance, when presented with direct evidence of some fact, such as that I have a bad leg, recognition of my intention is no longer a *reason* to come to believe some proposition, such as that I cannot play squash. For Sperber and Wilson this would be a case of ostensive-inferential communication; for Grice it would not be a case of non-natural meaning.

In other words, the Sperber and Wilson account can be understood as explaining the various ways to get others to believe certain things or behave in certain ways, including those where recognition of an intention is not necessary. Sometimes we do expect intention recognition (with MT utterances), and sometimes we display direct evidence (with ST), as captured by the Sperber and Wilson MT-ST continuum.[2]

1 For instance, Stephen Neale discusses dropping the third clause in his often-cited 1992 paper on Grice (Neale 1992). Most scholars acknowledge that whether or not this clause is needed can be a matter of what one is aiming to capture with their theory. As Sperber and Wilson note in the quotation on this page the third clause is likely needed for certain cases of meaning. For more on this see Johnson (2019).

2 In their 2015 paper Sperber and Wilson distinguish not only between meaning and showing but also between determinate and indeterminate content. For my purposes, I will be focusing on just determinate content, because otherwise the details of the theory become unwieldy. For further discussion of the Sperber and Wilson continua in all its complexity see Johnson (2019).

As a last bit of relevant theory before moving on to some motivating examples, let me also note that Sperber and Wilson present their account in terms of manifestness, understood as a technical term. When some content p is shown or meant, this is the sort of thing that makes p more manifest on the Sperber and Wilson picture. Manifestness is a combination of epistemic strength and salience. Manifestness is the extent to which, for any given proposition, the interlocutor "is likely to some positive degree to entertain it and accept it as true" (Sperber and Wilson, 2015, p. 134). 'Salience' here is what they called 'accessibility' in their original 1986 presentation in *Relevance* (Sperber and Wilson, 2015, p. 133). Manifestness is a property of the proposition itself, given the context.[3]

## Some motivating examples

Before moving on to the Fricker and social psychology literature discussion, let me now give 3 anecdotes that illustrate the difference between meaning that (MT) and showing that (ST). The first and second examples have me as the speaker, i.e., the one wishing to persuade in the exchange. In the third example I am the hearer, i.e., the one being persuaded in the exchange. I will first present the examples and then explain their relation to the Sperber and Wilson MT-ST continuum.

### Motivating example Case 1

Last spring I ordered a bracelet online. The package came on time as expected. I opened the sealed shipping box. There was no indication it had been opened. The shipping box contained a velvety bag, which contained a small padded box. The small padded box was empty. Strangely, it contained the price tag that should have been attached to the bracelet. The box had apparently not been tampered with so it seemed like the issue originated when it was packed. I wanted my bracelet or a refund for the money. I called the relevant customer service number and described the situation to them. I knew it sounded strange—because it was in fact strange. The person I spoke with on the phone asked me to send them a picture of the empty box and I did. They accepted this as satisfactory and sent me a new bracelet.

### Motivating example Case 2

Last summer I ordered 6 dresses online, persuaded by some huge end-of-season markdowns. I had a big event coming up and thought perhaps one of them would be suitable. I happened to be outside for the delivery and I accepted the box directly from the FedEx delivery man. When I got inside I noticed that the box was very squished. The original brown tape that sealed the top was opened and it had been haphazardly taped again with clear tape. I opened the box to find 2 of the 6 dresses inside.

My thinking here was shaped by my previous experience with the bracelet where I had been asked to send a photo. I saved the box—now I had some evidence that could show it had been opened and then resealed. I called the customer service number and explained the situation. They said there would be an "investigation." It did not sound promising. Next thing I knew I had a refund for the full purchase price to my credit card—so I ended up getting 2 dresses for free. I never needed the damaged box as evidence so I recycled it.

### Motivating example Case 3

A few years ago, I received an email from a student saying that she could not come to class because she had jury duty. Any professor is familiar with emails of this sort and we usually get multiple of them each week. My standard response, as I believe is the case for many other professors, is to let the student know I appreciate them reaching out and tell them that they should get the notes from a classmate and come to office hours if they have questions. If the student says they are sick I also tell them I hope they feel better soon. In almost all cases it does not matter to me if they are lying, and realistically I know a certain percentage will be. I emailed this standard response to the student who said she had jury duty. My student then replied again with a photo of her jury summons. I had not asked her for it.

## Discussion of cases and MT-ST continuum

In Case 1, I tried to get the customer service agent to believe that my bracelet had been missing from the package. I told her this verbally—a case of MT. This was not sufficient and she asked me to ST—provide "evidence"–and so I sent the photo of the empty box. In Case 2 I again tried to get the customer service agent to believe that 4

3 Other related work including (Sperber et al., 2010; Sperber and Mercier, 2012, 2017, 2018; Mercier, 2020) discusses epistemic vigilance, trust, and reason. For our purposes I will not be adopting this epistemic vigilance framework and instead adopt the manifestness notion from the Sperber and Wilson (2015). There are a number of reasons for this. Most importantly, because of the fact that Sperber and Wilson explicitly use the notion of manifestness in their 2015 paper it is clear how they see it working with the rest of their framework on showing vs. meaning. Beyond this, epistemic vigilance is presented as a state of the interpreter rather than of the proposition in a context. Sperber et al. (2010) and Sperber and Mercier (2017, 2018) claim that epistemic vigilance is an evolved mental module. They state that epistemic vigilance is "typically conscious" and "involves engaging in some higher order or metarepresentational thinking about one's own beliefs" (Sperber et al., 2010, p. 376). However, this sort of explicit reasoning on the part of an interpreter seems exceedingly rare. Seemingly aware of these issues, in their initial presentation of the notion of epistemic vigilance in 2010 the authors note that their paper included a number of assumptions "several of which we ourselves view as speculative" (Sperber et al., 2010, p. 384). Later formulations of this line of research (Sperber and Mercier, 2017, 2018) have not provided much needed clarity. Critics include Kim Sterelny, who noted that "appeal to a metarepresentational reasoning module seems not to help us at all" (Sterelny, 2018). Given such complications, I have chosen not to use the epistemic vigilance framework here. Again, because the notion of manifestness is a property of propositions in a context this requires positing no specific mental framework on the part of the interpreter.

of my items had been missing from the package. I told her this verbally—a case of MT. This was sufficient and I was not asked to ST—to send a photo of the box.

However, in Case 2 I still incurred the cost of saving the box. I was less confident that I would be believed just on the basis of my word. My expectation had been shaped by my previous experience where I first tried to get the agent to believe something with my word alone. Since this wasn't enough I prepared to show evidence in a similar interaction in the future.

In Case 3 the student first MT when she told me she had jury duty. The student ST when she sent the photo, providing me with direct evidence. I had not asked her for this photo. What exactly had caused her to make this shift? Did she send me the picture of her jury summons because she thought I believed she was lying? Just like me with the customer service agents, she likely had experienced a similar situation in the past. She likely had had a professor or teacher who did not accept her word as enough and asked for some sort of proof. She evidently thought that my response meant that I needed further documentation and thus provided it.[4]

As can be seen in Cases 1, 2, and 3, moving down the meaning-showing continuum can be a result of a speaker learning that recognition of her communicative intention has not in her experience been sufficient to induce the intended response in the hearer.

## Social interpretation

The costs of communication are not equal for all interlocutors—they may be greater for those who must show what they wish to make manifest to their hearers. Why would someone, on an occasion, choose to provide direct evidence in support of some fact rather than expect that their communicative intention alone would be enough to make some content manifest in the hearer? The answer has to do with how they expect they will be interpreted. If we reflect on social realities it becomes clear that manifestness as Sperber and Wilson define it—the extent to which, for any given proposition, an interlocutor "is likely to…accept it as true" (134) depends not just on the proposition itself but on *who says* that statement to us.

In presenting their account of relevance, Sperber and Wilson do hint at the role of power dynamics in their notion of optimal relevance to an individual[5] (Sperber and Wilson, 1986, p. 142–161). They write, "How much effort the addressee can expect the communicator to put into being relevant varies with the circumstances, the communicator, and the relationship between communicator and addressee" (Sperber and Wilson, 1986, p. 160). Later in the paragraph they illustrate this point noting that "A master talking to his servant may say whatever he wishes and merely assume that it will be relevant enough" (160). They also illustrate the point with an example of a woman named Mary who is to infer that she should make dinner when her surgeon husband says "I had a long day. I'm tired" (145–149). As they describe, when we engage in communication of this sort with well known interlocutors, we can bring an array of background assumptions to bear on the conversation. These mentions of a power dynamic do not receive further treatment but Sperber and Wilson are explicit to note that characterization of relevance to an individual is "psychologically more appropriate" (142).

These background assumptions about our interlocutors develop over time into what we might call a more or less refined "theory" about the speaker. Sometimes these "theories" are based on extensive knowledge of past interactions and other times they rely on rough heuristics. In the canonical paper "A Nice Derangement of Epitaphs" philosopher David Davidson draws our attention to such socially-relevant features of a speaker. As he notes an interpreter "alters his theory" about a speaker in light of these factors:

> An interpreter has, at any moment of a speech transaction, what I persist in calling a theory. (I call it a theory, as remarked before, only because a description of the interpreter's competence requires a recursive account.) I assume that the interpreter's theory has been adjusted to the evidence so far available to him: knowledge of the character, dress, role, sex, of the speaker, and whatever else has been gained by the speaker's behavior, linguistic or otherwise. As the speaker speaks his piece the interpreter alters his theory (Davidson, 2006, p. 260).

Let us take it as a given that Davidson has made an important point about the social realities of communication—which are often overlooked by philosophers of language. Davidson does not specify exactly how an interpreter would alter his theory in light of each of these factors, but we can now turn to philosopher Miranda Fricker to consider some specific relevant examples of just this very thing.

In her work Fricker presents a "socially situated account," which, again, she defines as "an account such that the participants are conceived not in abstraction from relations of social power…but as operating as social types who stand in relations of power to one another" (Fricker, 2006, p. 3). Fricker's account of the aims of testimony bears striking similarities to the Sperber and Wilson notion of manifestness and what Davidson discusses in the section just quoted above. Fricker explains,

> We are picturing hearers as confronted with the immediate task of gauging how likely it is that what a speaker has said is true.

4 Case 3 is unlike Cases 1 and 2 in that I of course do not have direct access to the minds of my students and thus am forced to speculate here. I will say that as someone on the receiving end of her communications I did stop and think about what I was doing that caused her to communicate in this way. My thought process can be explained in terms of the Principle of Relevance. Here it seems that the student surely wanted to communicate something further than what she had already achieved with the first email, given that it is "mutually manifest that the communicator intends it to be manifest to the addressee that she has chosen the most relevant stimulus capable of fulfilling her intentions" (Sperber and Wilson, 1986, p. 157) and that "to the best of the communicator's knowledge, the ostensive stimulus is relevant enough to be worth the audience's attention" (Sperber and Wilson, 1986, p. 156).

5 Although if that is what they had in mind with this case it is not made explicit. The terms 'race,' 'gender,' 'power,' 'class' are not found in the Appendix of Sperber and Wilson (1986).

> Barring a wealth of personal knowledge of the speaker as an individual, such a judgment of credibility must reflect some kind of social generalization about the epistemic trustworthiness—the competence and sincerity—of people of the speaker's social type, so that it is inevitable (and desirable) that the hearer should spontaneously avail himself of the relevant generalizations in the shorthand form of (reliable) stereotypes (Fricker, 2006, p. 32).

Gauging how likely it is that what a speaker has said is true in "face-to-face testimonial exchanges" requires a the hearer to, as Fricker writes, "make some attribution of *credibility* regarding the speaker. Such attributions are surely governed by no precise science, but clearly there can be error in the direction of excess or deficit" (Fricker, 2006, p. 18). Manifestness in the Sperber and Wilson sense clearly is not just a matter of the content of some proposition—it also depends who asserts this content to us. And it should: we should not take all people to be equally reliable sources of information, indiscriminately changing our beliefs regardless of who is the source. As Fricker writes, "Much of everyday testimony requires the hearer to engage in a social categorization of speakers, and this is how stereotypes oil the wheels of testimonial exchange" (Fricker p. 32). When faced with interpretive knowledge gaps we need to fill them in somehow.

To illustrate her points Fricker has us consider a case of the dependable family doctor (Fricker, 2006, p. 32). Consider the following utterance said by a family doctor:

"You will be at increased risk of heart attack if you get the new COVID-19 booster."

And consider again the utterance said by the person sitting next to you on the last airplane you took. We would likely give different weight to this utterance about COVID-19 boosters based on who said it.

Picture now very clearly that reliable family doctor. Get a fleshed-out mental picture. Consider now the gender, race, and accent of the family doctor you were picturing. What Fricker draws particular attention to in her work is the way that identity prejudice can be present in otherwise rational assessments of speaker credibility. She writes, "Many of the stereotypes of historically powerless groups such as women, black people, or working-class people variously involve an association with some attribute inversely related to competence or sincerity or both: over-emotionality, illogicality, inferior intelligence." (32) We do not fill in those gaps in the same way for all speakers.

Fricker calls "Testimonial Injustice" when "prejudice causes a hearer to give a deflated level of credibility to a speaker's word" (2007, p. 1). She vividly illustrates what Testimonial Injustice looks like with a discussion of Harper Lee's *To Kill a Mockingbird*.

> The year is 1935, and the scene a courtroom in Maycomb County, Alabama. The defendant is a young black man named Tom Robinson. He is charged with raping a white girl, Mayella Ewell, whose family's run-down house he passes every day on his way to work, situated as it is on the outskirts of town in the borderlands that divide where whites and blacks live. It is obvious to any reader, and to any relatively unprejudiced person in the courtroom, that Tom Robinson is entirely innocent. For Atticus Finch, our politely spoken counsel for the defense, has proved beyond doubt that Robinson could not have beaten the Ewell girl so as to cause the sorts of cuts and bruises she sustained that day, since whoever gave her the beating led with his left fist, whereas Tom Robinson's left arm is disabled, having been injured in a machinery accident when he was a boy. The trial proceedings enact what is in one sense a straightforward struggle between the power of evidence and the power of racial prejudice, with the all-white jury's judgment ultimately succumbing to the latter (Fricker, 2006, p. 23).

Fricker presents this case as a "struggle between the power of evidence and the power of racial prejudice." We also see illustrated in this case a struggle between MT and ST. In claiming that Tom Robinson raped her Mayella Ewell is able to MT and be believed. Tom Robinson, through his lawyer Atticus Finch, knows that to simply MT in reply will not lead the jury to believe that Tom is innocent. He must provide direct evidence.

Lawyers in presenting their cases do sometimes rely on MT. They coach witnesses on how to appear credible and bring in experts (Loftus and Ketcham, 1992; Elm, 2008). But in Tom's case—given how he will be perceived as an African American man at this time in America—ST is needed. Atticus in representing his client moves down the MT-ST continuum. As readers who know his innocence we hope this will be enough. But it still is not. As Fricker writes, "They fail, as Atticus Finch feared, precisely in their duty to believe Tom Robinson" (26).

Of course, most situations in which we try to convince someone of some proposition are not played out in the court of law, but in more informal circumstances. We do see parallels however in "the court of the professor's decision" and "the court of the customer service representative." Depending on the stereotypes we have about a speaker they will sometimes be able to persuade with MT, sometimes with ramping things up to ST, and sometimes not even ST will be enough.

Fricker's observations, as she notes, are borne out not just by fictions such as *To Kill a Mockingbird*, but by social psychology research as well. Fricker cites psychologist Taylor (1982) who writes, "Empirical work on non-social judgments indicates that the perceiver employs shortcuts or heuristics to free capacity and transmit information as quickly as possible." Fricker notes that this need not be conscious or deliberate, citing Kahneman and Tversy (1973), whose work on System 1 and System 2 shows a number of the mental shortcuts that we make every day, and the ways they are subject to systematic and predictable errors (Kahneman, 2013). For instance, after being presented with an anchor of some number, participants when then asked to estimate some quantity are more likely to give a figure closer to that anchor than those who have not been primed in this way (Kahneman and Tversy, 1973; Kahneman, 2013).

Again, circling back to the quotation by Davidson, not all speakers are perceived in the same way. The speaker's "the character, dress, role, sex, of the speaker, and whatever else has been gained by the speaker's behavior, linguistic or otherwise" (Davidson, 2006) can serve as a sort of "anchor" for how much credibility they are afforded by a hearer.

Accent is one of the clearest ways that credibility can be affected in the eyes of the interpreter, and there has been much empirical research conducted on this topic (Dixon et al., 2002; Kinzler et al., 2007; Lev-Ari and Keysar, 2010; Dragojevic et al., 2021). There are a number of ways that accents can be classified and they can signal class, race, gender, country of origin and many other things. One way that researchers Dragojevic et al. classified accents in a recent summary paper on one hundred years of language research is as "low prestige"

and as "high prestige" (Dragojevic et al., 2021). They write that, "Research shows that language varieties within a given society can be ordered on a hierarchy of prestige, typically corresponding to the socioeconomic status of the social groups they are associated with. Varieties associated with socioeconomically dominant groups tend to carry high prestige; these typically include majority group languages, standard varieties—namely those that have been codified" (Dragojevic et al., 2021, p. 63). They define low prestige varieties of language as those we "associate with socioeconomically subordinate groups," and continue to note that "these typically include minority group language, nonstandard varieties—namely those that diverge from codified norms, including most regional and ethnic dialects and foreign accents—and other forms linked to stigmatized groups (e.g., gay/lesbian speech)" (Dragojevic et al., 2021, p. 63). As we might expect, their summary of one hundred years of language research shows that "Speakers of low prestige varieties [of language] frequently face prejudice and discrimination" (Dragojevic et al., 2021, p. 67). This is just the sort of thing Fricker draws our attention to in her work.

Further, this bias against low prestige varieties of language is found even earlier than one might expect. As Dragojevic et al. note, at 5 months infants can distinguish between native and foreign accents and "express a clear social preference for native- over foreign-language speakers, without any knowledge of specific linguistic stereotypes" (Kinzler et al., 2007; Dragojevic et al., 2021, p. 63). At age 10 to 12 months infants are "more likely to accept toys from native over foreign language speakers" and preschoolers trust native-language friends more than foreign-language friends (Kinzler et al., 2007).

One place where older children might learn this bias against non-native speakers is in the media, if they are not already exposed to it in their everyday life. For, in an analysis of Disney movies, cartoon, and primetime television it was found that "standard speakers tend to be portrayed in positive roles, whereas nonstandard speakers—particularly foreign-accented speakers—in negative roles" (Dragojevic et al., 2021, p. 67). From childhood we are conditioned to trust certain speakers less than others. Unsurprisingly, this carries over into adulthood, where nonstandard speakers "tend to be judged as less credible, truthful, and accurate eyewitnesses[6]" (Dragojevic et al., 2021, p. 68). Accent cues are just one of the many ways that an interlocutor can have a "credibility excess" or "credibility deficit."

There are many other cues present in speech and bodies including but not limited to the perception of the speaker's race, gender, class, and age. In addition, a speaker's vocal pitch and pacing affect how they are perceived. Vocal pitch is associated with size in humans and animals (Sell et al., 2010). Empirical research has demonstrated that vocal pitch in both men and women is correlated with perception of leadership quality, attractiveness, and strength (Zuckerman and Miyake, 1993; Sell et al., 2010; Klofstad et al., 2012). In a study on pitch and politics the authors conclude that "because women, on average, have higher-pitched voices than men, voice pitch could be a factor that contributes to fewer women holding leadership roles" (Klofstad et al., 2012). Acoustic analysis has demonstrated that certain linguistic features are associated with trustworthiness, independent of attribution of a gender to the speaker, including "accelerated tempo, low harmonic-to-noise ratio, more shimmer, low fundamental frequency, more jitter, large intensity range" (Schirmer et al., 2020). Physical bodily features also affect how a "credibility excess" or "credibility deficit" is attributed to interlocutors. Those who are "babyfaced" are thought to be less competent (Zebrowitz and Montepare, 2005). Taller people are thought to be more natural leaders and earn more money (Judge and Cable, 2004; Maclean, 2019). The taller candidate has won the U.S. election two-thirds of the time (Maclean, 2019).

When speakers are perceived to have certain features of any kind that give them a credibility deficit, it is rational to do a number of things to lessen these appearances. These can include working to change one's accent or style of dress.

These changes made to be perceived as a different sort of interlocutor are prevalent. We see this reflected in fictions such as Eliza Doolittle in the famous Shaw play *Pygmalion*, and with characters such as Lucien de Rubempré in Balzac's *Lost Illusions*. These fictions ring true because they capture a reality that persists today.

Given the knowledge that how we appear changes how likely we are to be believed, it is only rational to make changes to be perceived more favorably. Being overweight is associated with being poor and uneducated and thus it is "economically rational for ambitious women to try as hard as possible to be thin" (Economist, 2022). Adorning the body in a way that changes the perception of the physical body and thus the associated meanings is what I have called in other work imitation of natural meaning (Johnson, 2022a), drawing here on Grice's distinction between natural and non-natural meaning. This has been seen throughout history and can happen in ways large and small, from dying one's hair, to wearing a suit, to wearing makeup (Johnson, 2022a). Female politicians are coached so as to appear feminine to the right degree, down to details like changing what they wear, being coached on the pitch of their voice, and how to reduce small gestures such as touching their hair, which are perceived negatively (Jahnke, 2011). Attorneys, too, pay great sums of money to jury consultants who coach them on how they are perceived—often leading to feedback that is deemed "superficial" by attorneys but which substantially affects jury rulings, such as that an attorney needs to smile more or less (Kressel and Kressel, 2004, p. 4; Postal, 2022).

This is not to say that these changes are without costs—financial as well as emotional—or that this is the way that things ought to be. Many of these efforts to reduce an appearance that lead to a credibility deficit will be quite taxing (Du Bois, 1903/2016; Jahnke, 2011; McCluney et al., 2019), and reflect the sexism, racism, ableism, and other of the worst biases in our society. However, as individuals it is often wise to act prudentially with how we present ourselves to the world [distinguishing this from the times where we have a moral obligation to resist these stereotypes (Jeffers, 2012; Cray, 2021; Johnson, 2022b)]. On top of this, such efforts are unlikely to be entirely effective—presenting one's self in certain ways can lessen the effect of the credibility deficit but it will usually not be fully eradicated. It also can lead to other costs: charges of being a traitor to one's community (Barnes, 2022).

Speakers who are perceived to have a credibility deficit may move down the MT-ST continuum, as we have seen. I found myself doing this with what I learned between Case 1 and 2, and speculated that it is what motivated my student in Case 3. However, this is not the only

6 They write, "In simulated criminal proceedings, nonstandard speakers are often judged as more guilty than standard speakers…and as more likely to be re-accused of a crime, regardless of the quality of the evidence presented against them" (Dragojevic et al., 2021, p. 68).

change it would be rational for them to make. Recall that manifestness is an explicitly epistemic notion, the extent to which, for any given proposition, the interlocutor "is likely to some positive degree to entertain it and accept it as true" (Sperber and Wilson, 2015, p. 134). Through the discussion I have presented here we see two specific types of response to Testimonial Injustice emerge—the first understood as a move down the MT-ST continuum and the second understood as the steps taken to be perceived as more credible in the first place, to increase the likelihood that one's MT will be enough. These are both a result of a speaker learning that recognition of her intention has not in her experience been sufficient to induce an intended response in the hearer.

We see the following two categories emerge:

1. **"Prove it" or Showing-That Injustice** – This is when a communicator expends extra time and resources presenting an interlocutor with direct evidence for some proposition that—barring identity prejudice—they would accept without such direct evidence.
2. **"Look it" or Personal-Appearance-Modification Injustice** – This is when a communicator expends extra time and resources presenting themselves in a way that makes them be perceived in a way that lessens their credibility deficit. This includes all forms of changes in adornment and bodily styling, as well as changes to accent, vocabulary, and manner of speech.

On the part of the speaker[7] Showing-That and Personal-Appearance-Modification Injustice are rational response to past Testimonial Injustice, and includes both the intentional as well as automatic, unconscious changes.

## Takeaways

Many explanations of why we engage in communicative acts attempt to account for the cost of communication—an assumption that underpins Sperber and Wilson's presumption of relevance.[8] The potential for Fricker's epistemic injustice theory to be applied directly to philosophy of language is made evident by Sperber and Wilson's framework, as well as by their clear spelling out of manifestness as an explicitly *epistemic* notion. One of the types of moves we automatically take in processing information is to use stereotypes and heuristics of the sort that Fricker draws attention to in her work. This leads to discrepancies in the effort that different types of speakers have to expend in making their meanings manifest to interpreters. This leads to further forms of injustice because when certain speakers systematically face a credibility deficit they must expend more resources to be believed. This extends to the actions taken before making an utterance, as well as those that follow the recognition that a hearer requires direct evidence.[9]

In thinking about philosophy of language, we should be asking not only *how* but *why* we engage in certain forms of communicative behavior—such as why do we sometimes show and other times mean content. Often the answer to these types of questions lies in the details of social factors, of the sort that philosophers often gloss over in developing their theories. We should not ignore how questions of meaning and interpretation are shaped by the power dynamics at play between interlocutors. There is a time for abstraction; and, there is a time for addressing these social realities of communication.

## Author's note

This paper has benefited enormously from the feedback of others. Thank you to the audience at the Philosophy of Linguistics and Language 2022 conference in Dubrovnik, especially Josh Armstrong, Michael Glanzberg, Stephen Neale, Peter Pagin, David Sosa, and Una Stojnić, as well as the reviewers for this journal for their detailed comments and suggestions. This paper would not have been written without the kind invitation for this Research Topic so I would like to thank the editors for their organizing efforts and support of the paper.

## Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

## 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.

## Publisher's note



7 Although I focus on speakers here, in future work I hope to explore how these forms of injustice are also prevalent in those who want to be perceived as more credible hearers.

8 They write, "The key problem for efficient short-term information processing is to achieve an optimal allocation of central processing resources. Resources have to be allocated to the processing of information which is likely to bring about the greatest contribution to the mind's general cognitive goals at the smallest processing cost…Our claim is that all human beings automatically aim at the most efficient information processing possible" (Sperber and Wilson, 1986, p. 48–49).

9 Relevance theory is a hearer-focused account but of course hearers are not operating in isolation from the speaker. Sperber and Wilson discuss the ways the relationship between hearers and speakers shapes relevance in context and consider the ways that speakers aim their utterances, and sometimes fail. e.g., see their discussion of bores on page 158 and the section on masters and servants gauging relevance in a context quoted earlier.

## References


Barnes, C. (2022). An account of normative stereotyping. *J. Ethics Soc. Philos.* 21, 396–421. doi: 10.26556/jesp.v21i3.1384

Carston, R. (2005). Relevance theory, Grice, and the neo-Griceans: a response to Laurence Horn's 'current issues in neo-Gricean pragmatics. *Intercult. Pragmat.* 2, 303–319. doi: 10.1515/iprg.2005.2.3.303

Cray, L. (2021). Some considerations regarding adornment, the gender 'binary', and gender expression. *J. Aesthetics Art Critic.* 79, 488–492. doi: 10.1093/jaac/kpab051

Davidson, D. (2006). "A nice derangement of epitaphs" in *The essential davidson*. eds. E. Lepore and K. Ludvig (Oxford: Oxford University Press)

Dixon, J. A., Mahoney, B., and Cocks, R. (2002). Accents of guilt? Effects of regional accent, 'race' and crime type on attributions of guilt. *J. Lang. Soc. Psychol.* 21, 162–168. doi: 10.1177/02627X02021002004

Dragojevic, M., Fasoli, F., Cramer, J., and Rakić, T. (2021). Toward a century of language attitudes research: looking back and moving forward. *J. Lang. Soc. Psychol.* 40, 60–79. doi: 10.1177/0261927X20966714

Du Bois, W. E. B. (1903/2016). *The souls of black folk*. New York: Pocket Book.

Economist (2022). "*The economics of thinness: it is economically rational for ambitious women to try as hard as possible to be thin*".

Elm, D. L. (2008). *Vouching: a defense Attorney's guide to witness credibility, law, and strategy*. Chicago, IL: American Bar Association Book Publishing.

Fricker, M. (2006). *Epistemic injustice*. New York, NY: Oxford University Press.

Horn, L. (2006). More issues in neo- and post-Gricean pragmatics. *Intercult. Pragmat.* 3, 81–93. doi: 10.1515/IP.2006.004

Jahnke, C. K. (2011). *The well-spoken woman: your guide to looking and sounding your best*. Lanham, MD: Prometheus.

Jeffers, C. (2012). "Should black kids avoid wearing hoodies?" in *Pursuing trayvon martin: historical contexts and contemporary manifestations*. eds. G. Yancy and J. Jones (Lanham: Lexington Books), 129–140.

Johnson, M. (2019). Making meaning manifest. *Croatian J. Philos.* 19, 497–520. doi: 10.5840/croatjphil201919329

Johnson, M. (2022a). *Adorning bodies: meaning, evolution, and beauty in humans and animals*. London: Bloomsbury Academic.

Johnson, M. (2022b). The meaning of bodies. *Philos. Mag.* 4, 66–73. doi: 10.5840/tpm20229867

Judge, T., and Cable, D. (2004). The effect of physical height on workplace success and income: preliminary test of a theoretical model. *J. Appl. Psychol.* 89, 428–441. doi: 10.1037/0021-9010.89.3.428

Kahneman, D. (2013). *Thinking fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Kahneman, D., and Tversy, A. (1973). Availability: a heuristic for judging frequency and probability. *Cogn. Psychol.* 5, 207–232.

Kinzler, K. D., Dupoux, E., and Spelke, E. S. (2007). The native language of social cognition. *Natl. Proc. Nat. Acad. Sci. U. S. A* 104, 12577–12580. doi: 10.1073/pnas.0705345104

Klofstad, C., Anderson, R. C., and Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proc. R. Soc. B* 279, 2698–2704. doi: 10.1098/rspb.2012.0311

Kressel, N., and Kressel, D. (2004). *Stack and Sway: The New Science of Jury Consulting*. Boulder, CO: Westview Press.

Lev-Ari, S., and Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *J. Exp. Soc. Psychol.* 46, 1093–1096. doi: 10.1016/j.jesp.2010.05.025

Loftus, E., and Ketcham, K. (1992). *Witness for the defense: the accused, the eyewitness, and the expert who puts memory on trial*. New York: St. Martin's Griffin.

Maclean, D. (2019). "American tend to elect the tallest person for president". The Independent. November 18. Accessed online.

McCluney, C. L., Robotham, K., Lee, S., Smith, R., and Durkee, M. (2019). The costs of code-switching: the behavior is necessary for advancement – but it takes a great psychological toll. *Harvard Bus Rev*.

Mercier, H. (2020). *Not born yesterday: the science of who we trust and what we believe*. Princeton: Princeton University Press.

Neale, S. (1992). Paul Grice and the philosophy of language. *Linguist. Philos.* 15, 509–559. doi: 10.1007/BF00630629

Postal, K. (2022). *The Art and Science of Expert Witness Testimony: A Multidisciplinary Guide for Professionals*. New York, NY: Routledge.

Schirmer, A., Chiu, M. H., Lo, C., Feng, Y.-J., and Penney, T. B. (2020). Angry, old, male – and trustworthy? How expressive and person voice characteristics shape listener trust. *PLos One* 15:e0232431. doi: 10.1371/journal.pone.0232431

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., et al. (2010). Adaptations in humans for assessing physical strength from the voice. *Proc. R. Soc. B* 277, 3509–3518. doi: 10.1098/rspb.2010.0769

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., et al. (2010). Epistemic Vigilance. *Mind Lang.* 25, 359–393. doi: 10.1111/j.1468-0017.2010.01394.x

Sperber, D., and Mercier, H. (2012). "Reasoning as social competence" in *Collective wisdom*. eds. J. Elster and H. Landemore (Cambridge: Cambridge University Press), 368–392.

Sperber, D., and Mercier, H. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.

Sperber, D., and Mercier, H. (2018). Why a modular approach to reason? *Mind Lang.* 33, 533–541. doi: 10.1111/mila.12208

Sperber, D., and Wilson, D. (1986). *Relevance: communication and cognition*. Oxford: Wiley-Blackwell.

Sperber, D., and Wilson, D. (2015). Beyond speaker's meaning. *Croatian J. Philos.* 15, 117–149. doi: 10.5840/croatjphil20151528

Sterelny, K. (2018). Why reason? Hugo Mercier's and Dan Sperber's the enigma of reason: a new theory of human understanding. *Mind Lang.* 33, 502–512. doi: 10.1111/mila.12182

Taylor, S. E. (1982). "The Availability Bias in Social Perception and Interaction" in *Judgment Under Uncertainty: Heuristics and Biases*. (eds.) D. Kahneman, P. Slovic and A. Tversky Cambridge: Cambridge University Press, 190–200.

Zebrowitz, L. A., and Montepare, J. M. (2005). Appearance DOES matter. *Science* 308, 1565–1566. doi: 10.1126/science.1114170

Zuckerman, M., and Miyake, K. (1993). The attractive voice: what makes it so? *J. Nonverbal Behav.* 17, 119–135. doi: 10.1007/BF01001960

# Frontiers in Psychology

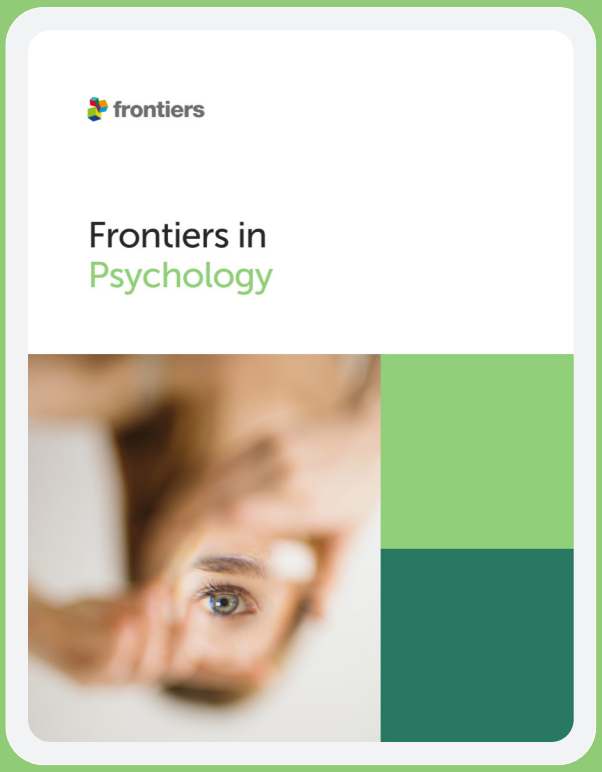
**Paving the way for a greater understanding of human behavior**

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

---

## Discover the latest Research Topics

See more →




**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact