GESTURE-SPEECH INTEGRATION: COMBINING GESTURE AND SPEECH TO CREATE UNDERSTANDING

EDITED BY: Naomi Sweller, Kazuki Sekine and Autumn Hostetter

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-312-7
DOI 10.3389/978-2-88971-312-7

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

GESTURE-SPEECH INTEGRATION: COMBINING GESTURE AND SPEECH TO CREATE UNDERSTANDING

Topic Editors:
Naomi Sweller, Macquarie University, Australia
Kazuki Sekine, Waseda University, Japan
Autumn Hostetter, Kalamazoo College, United States

Citation: Sweller, N., Sekine, K., Hostetter, A., eds. (2021). Gesture-Speech Integration: Combining Gesture and Speech to Create Understanding. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-312-7

Table of Contents

EDITORIAL
published: 23 July 2021
doi: 10.3389/fpsyg.2021.732357

Editorial: Gesture-Speech Integration: Combining Gesture and Speech to Create Understanding

Naomi Sweller[1]*, *Kazuki Sekine*[2] *and Autumn B. Hostetter*[3]

[1] *Department of Psychology, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW, Australia,* [2] *Department of Human Informatics and Cognitive Science, Faculty of Human Sciences, Waseda University, Saitama, Japan,* [3] *Department of Psychology, Kalamazoo College, Kalamazoo, MI, United States*

Keywords: gesture, semantic meaning, comprehension, communication, cognitive load, individual differences, language

Editorial on the Research Topic

Gesture-Speech Integration: Combining Gesture and Speech to Create Understanding

Gestures and speech are tightly linked. Since McNeill (1992) argued that gesture and speech form a single integrated system, research has shown that gestures and speech interact with each other across a variety of domains. Listeners can benefit from observing a speaker's gestures (e.g., Kelly, 2001), and similarly, speakers demonstrate improved communication and task performance when they gesture (e.g., Cook et al., 2010). In 15 articles, this Special Issue further examines how gesture and speech are integrated during speaking and listening. The functions of gesture, and potential mechanisms underlying gesture's beneficial effects are considered, and together, these articles highlight the impact that both producing and observing gestures can have on individuals' learning and communication across the lifespan. Here, we summarize some of the overarching themes that emerge from this collection.

Gesture seems to activate semantic meanings that are useful for comprehension and learning. Hughes-Berheim et al. found that participants' ratings of the semantic congruency of gesture-word pairs were similar, regardless of whether the word was presented in speech or in text. This suggests that gestures activate semantic meanings that are independent of language modality. Further, the meanings conveyed by gesture are particularly helpful for children's learning. Guarino and Wakefield examined 4–11-year-old children's understanding of instructions presented through speech alone, or through a combination of speech and gesture. They found a benefit of the combination of gesture and speech beyond speech alone that was most marked for 5-year-old children. Eye-tracking results suggested that the gestures may have helped children to organize their attention and clarify ambiguous spoken instructions. In addition to these attention-related functions, the semantic meaning activated by gesture can act as a cue during retrieval to help children remember what they learned. Mertens and Rohlfing compared progressively reduced iconic gestures with fully executed iconic gestures during children's recall of words. Although children's recall of the target words was unaffected by the type of gesture observed, their production of the target words at test was enhanced by progressively reduced gestures relative to fully executed gestures.

By activating semantic meaning, gestures help speakers and listeners resolve ambiguous references. Debreslioska and Gullberg examined the relationship between the information status of a referent (brand-new vs. inferable referants) and gesture, finding that gestures were more frequent with inferable than with brand-new referents. This finding suggests a function of gestures for disambiguating discourse content. Hinnell and Parrill found that listeners relied on a speaker's gesture as an indication of what the speaker's own opinion was. Speakers presented two contrasting

OPEN ACCESS

Edited and reviewed by:
Bernhard Hommel,
University Hospital Carl Gustav Carus, Germany

****Correspondence:***
Naomi Sweller
naomi.sweller@mq.edu.au

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

Received: *29 June 2021*
Accepted: *29 June 2021*
Published: *23 July 2021*

Citation:
Sweller N, Sekine K and Hostetter AB (2021) Editorial: Gesture-Speech Integration: Combining Gesture and Speech to Create Understanding. Front. Psychol. 12:732357. doi: 10.3389/fpsyg.2021.732357

ideas, and then said which one they agreed with. When the speaker accompanied their spoken agreement with a gesture, listeners were more likely to state that the speaker preferred the idea accompanied by the gesture. In this way, gestures activate semantic meaning that helps listeners infer what is meant when speech is not completely clear.

Even without accompanying speech, the semantic meanings conveyed by gesture are important for communication. Marentette et al. showed this in children's production of pantomime gestures, or non-co-speech gestures, that were performed during children's spoken narratives. Marentette et al. found that narratives that included non-co-speech gestures were longer and sometimes of higher quality than those with only co-speech gestures, suggesting that expressing information uniquely in gesture (and not in speech) can improve the overall quality of children's narratives. Hsu et al. make a similar point based on their analysis of gestures taken from a corpus of American TV talk shows. They discuss many examples of what they call "speech-embedded nonverbal depictions," that is, non-verbal communicative cues presented iconically, but without simultaneously co-occurring speech. The authors argue that such depictions are frequently overlooked in the literature, and argue for their theoretical and functional significance. Taken together, these papers demonstrate how gestures activate semantic meanings that do not rely on accompanying speech and that contribute to the on-going narrative.

The benefits of gesture for comprehension also go beyond the purely semantic; gestures can also affect other areas of language processing from low-level phonemic recognition to high-level social judgments about the speaker. Hoetjes and Maastricht examined second language (L2) phoneme acquisition, with a focus on the complexity of both the phonemes and of the gestures observed. Gestures were either simple (pointing) or more complex (iconic) gestures, and the to-be-learned Spanish phonemes were either simple (contained in the Dutch phoneme inventory) or complex (not contained in the Dutch phoneme inventory). While the more complex gesture enhanced learning of the simple phoneme, it was detrimental to learning the complex phoneme. At the other end of the spectrum, gestures can also affect high-level social judgments about a speaker. Billot-Vasquez et al. found that native Mandarin and Japanese speakers evaluated the accents of non-native speakers and the non-native speakers themselves more favorably when they produced a familiar emblematic gesture with their speech compared to producing the speech alone. These papers suggest that gestures can contribute more to language than just activating a particular semantic meaning.

Even as gestures have these positive effects, they may also come with costs in some situations. Specifically, producing or processing a gesture may impose an additional cognitive cost for some speakers and listeners. This was shown by Rohrer et al. in the case of beat gestures (rhythmic hand movements without any semantic meaning) that accompanied speech in a listener's non-native language. Specifically, French intermediate learners of English watched a video of a speaker describing a short narrative event in either French or English using either beat gestures or no gesture. When the learners drew a depiction of the narrative, it was found that recall of the narrative was negatively affected by beat gestures when the narrative was presented in their non-native language. The authors propose that these gestures may have increased cognitive load. Further, Overoye and Wilson examined gesture's effects on working memory load during a verbal reasoning task. Gesturing while explaining verbal analogies did not alleviate the load on working memory (as has been shown in previous studies—e.g., Goldin-Meadow et al., 2001), but rather led to poorer performance on a secondary task than being prohibited from gesturing.

How can gestures simultaneously be helpful in some ways and detrimental in others? One possibility is that it depends on the speakers' or listeners' cognitive skillset. Such is the suggestion in Özer and Göksun's timely review article, in which they explore how individual differences in cognitive capacity might affect people's gesture production, and the extent to which they employ gesture as a tool for comprehension. Özer and Göksun conclude that gestures can be used as a tool to compensate for a lack of cognitive resources, by both speakers and listeners. Indeed, it is well-recognized that speakers' gestures are affected by individual differences, including cognitive skills and also neurodevelopmental factors. For example, Huang et al. discuss how the gestures produced by Chinese-speaking children on the autism spectrum differ from those of their typically developing peers.

The potential for gesture production to differ depending on a speaker's cognitive skillset is further explored in Gordon and Ramani's new model, which integrates the information processing approach to children's mathematical problem solving with the theory of embodied cognition, frequently used in gesture studies. While the model does not differentiate between speech and gesture input, it does differentiate between the gestures and speech that children produce: even with similar speech output, individual differences in math knowledge are proposed to affect children's gesture production.

Finally, the fact that findings about the benefit of gesture often conflict across studies is highlighted in the review article by Arachchige et al. The authors note methodological variations across the field and discuss how these differences may contribute to the heterogeneity of findings, limiting our ability to draw conclusions regarding underlying mechanisms.

Together, these articles demonstrate the critical role that gesture holds in human cognition and communication. Whether we are producing gestures ourselves, or observing those performed by others, gestures and speech interact in profound, and sometimes unexpected, ways. Gestures can aid comprehension and learning through semantic links with speech, and can have a similarly important role in the absence of speech. Gestures can affect social evaluations of speakers, but can sometimes come with associated cognitive costs. The effects of gesture must be examined in the context of individuals' cognitive characteristics, as well as differences in the gestures themselves. The articles in this collection further our understanding of human communication, highlighting the range of tasks, ages, individual differences and methods through which we may examine the integration of gesture and speech.

REFERENCES

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2010). Gesturing makes memories that last. *J. Mem. Lang.* 63, 465–475. doi: 10.1016/j.jml.2010.07.002

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395

Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *J. Child Lang.* 28, 325–349. doi: 10.1017/S0305000901004664

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: The University of Chicago Press.

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.

AUTHOR CONTRIBUTIONS

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

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.

BRIEF RESEARCH REPORT
published: 08 September 2020
doi: 10.3389/fpsyg.2020.573212

Gestures in Storytelling by Preschool Chinese-Speaking Children With and Without Autism

Ying Huang, Miranda Kit-Yi Wong, Wan-Yi Lam, Chun-Ho Cheng and Wing-Chee So*

Department of Educational Psychology, The Chinese University of Hong Kong, Hong Kong, China

Previous findings on gestural impairment in autism are inconsistent, while scant evidence came from Chinese-speaking individuals. In the present study, preschool Chinese-speaking children with typical development and with autism were asked to generate stories from a set of wordless Cartoon pictures. Two groups were matched in chronological age and language developmental age. Their speech and gestures were coded. Compared to children with typical development, children with autism produced fewer gestures and showed lower gesture rate. Besides, children with autism produced fewer emblems and fewer supplementary gestures compared to their TD peers. Unlike children with typical development, children with autism tend to produce emblems for reinforcing, rather than supplementing information not conveyed in speech. Results showed the impairments in integrating the cross-modal semantic information in children with autism.

Keywords: gesture, autism spectrum disorder, storytelling, emblem, supplementary relation

OPEN ACCESS

Edited by:
Naomi Sweller,
Macquarie University, Australia

Reviewed by:
Laura M. Morett,
University of Alabama, United States
Ric Dalla Volta,
University of Catanzaro, Italy

****Correspondence:***
Ying Huang
huangyapril@gmail.com

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

Received: *16 June 2020*
Accepted: *14 August 2020*
Published: *08 September 2020*

Citation:
Huang Y, Wong MK-Y, Lam W-Y, Cheng C-H and So W-C (2020) Gestures in Storytelling by Preschool Chinese-Speaking Children With and Without Autism. Front. Psychol. 11:573212. doi: 10.3389/fpsyg.2020.573212

INTRODUCTION

Children typically exhibit communicative behaviors during their first year. Although spoken words become a preferred form of communication after the first year of development, children continue to gesture to reinforce or extend spoken messages or even to replace them (Colletta and Guidetti, 2012). Gestures refer to actions that are made with the intention of communication, and they can involve the hands, the fingers, and the whole body (Bochner and Jones, 2003). Gestural skills are crucial for facilitating communication. Gestures provide semantic information in a visual format (Goldin-Meadow, 2006) and help listeners understand speech better, especially when the co-occurring speech underspecifies information (Hostetter, 2011).

In comparison to children without autism spectrum disorder (ASD), children with ASD have a delay in verbal and nonverbal communication skills (Lai et al., 2014). Most of the children diagnosed with autism disorder show a significant delay in language development (Tager-Flusberg et al., 2005). Impairments in nonverbal skills, such as the use of gestures, from early childhood to school age in children with ASD have also been reported. In comparison to their typically developing (TD) peers, children with ASD have deficits in understanding and producing gestures (Colgan et al., 2006; Mitchell et al., 2006). Preschool children with ASD rarely use deictic gestures (i.e., pointing) to attract others' attention or to share their interest with others (Camaioni et al., 2003). Compared to mentally retarded children matched on mental age or language age, children with ASD showed deficits in gestural joint attention skills, which predicted their language development (Mundy et al., 1990). They have difficulties in understanding and producing conventional gestures (also known as emblems), such as waving hands to represent "goodbye" (Stone et al., 1997; Wetherby et al., 1998). In addition, the production of iconic and beat gestures is delayed in children with ASD (Charman et al., 2003; Wetherby et al., 2004; Luyster et al., 2007). It is also found that children with ASD imitate gestures worse than TD children and are more likely to make errors (Smith, 1998). Moreover, school-aged children with ASD are less able to perceive and produce gestures (Schreibman et al., 2015; So et al., 2015). It was reported that adolescents with ASD produce fewer metaphorical and beat gestures than their TD peers with matched age and verbal IQ (Morett et al., 2016). Researchers have argued that although verbally fluent teenagers use the same type of gestures as their TD peers, their gestures are more difficult to understand (Silverman et al., 2017).

However, evidence of impairment in the use of gestures is inconsistent across studies. For example, Mastrogiuseppe et al. (2015) reported that the amount of gestures produced by children with ASD is significantly lower than TD children and children with Down Syndrome. Conversely, Wong and So (2018) found that, compared to TD children, children with ASD produce a similar number of pointing gestures and markers and more iconic gestures. Similarly, de Marchena et al. (2019) reported that adults with autism used gestures more than TD controls for regulating conversational dynamics. When examining gesture rates (number of gestures per utterance), some researchers reported lower gestures rates in the ASD group (So et al., 2015; Morett et al., 2016; Silverman et al., 2017) while others found comparable gesture rates between the ASD group and the TD group (de Marchena and Eigsti, 2010). Similarly, findings on gesture types also vary. So et al. (2015) found that children with ASD use fewer types of gestures, while Silverman et al. (2017) suggest that the proportion of gesture types did not differ between the ASD group and TD group. In regard to gesture quality and meaning, Morett et al. (2016) and So et al. (2015) reported fewer, or even the absence of, supplementary gestures in children with ASD. However, Wong and So (2018) found that children with ASD produced comparable supplementary gestures to TD children. Moreover, the use of gestures may vary across cultures (Kita, 2009).

Most of the previous studies on gestural skills in individuals with ASD are based on English-speaking participants. Some recent studies reporting delayed and deficit in gestural use in school-aged Chinese-speaking participants with ASD (So et al., 2015, 2016, 2018). These results suggesting that early intervention is critical. However, little is known about the use of gestures and gestural skills in preschool Chinese-speaking children with or without ASD. This study examined whether Chinese-speaking children with ASD had impairments in gestural production skills compared to their age-matched TD peers. A narrative elicitation task was used to assess the rates, types, and gesture-speech relation produced spontaneously during storytelling. We expected that results would be consistent with previous findings of So et al. (2015). Specifically, we expected that children with ASD would produce fewer gestures, especially fewer emblems, and fewer supplementary gestures than their age-matched TD peers. Results could provide evidence for designing effective intervention programs.

METHODS

Participants

Twenty children with ASD and 14 TD children participated in the current study. All participants were native speakers of Chinese (Cantonese) aged 4 to 6. Children in the ASD group had been diagnosed with autism or autistic disorder by pediatricians at the Child Assessment Center for the Department of Health in Hong Kong. All procedures in the present study were approved by the institutional review board of the author's university, in compliance with the Declaration of Helsinki (Reference no. 14600817). Before the study, we explained the procedures to the parents and obtained their approval for videotaping. The participants also gave their assent to participate in the study.

The mean chronological age of the participants was 5.60 years (*SD* = 0.70; range 4.6–6.7) in the TD group and 5.51 years (*SD* = 0.44; range 4.7–6.3) in the ASD group, Mann-Whitney (*U*) = 127, *p* = 0.66. There was no significant difference between participants with ASD and those with typical development. Participants' language developmental age was assessed by the language and communication subset in Psychoeducational Profile, Third Edition (PEP-3; Schopler et al., 2005). Trained experimenters gave instructions in Chinese (Cantonese), which followed the Chinese version of PEP-3 (Shek and Yu, 2014). The mean language developmental age of participants was 5.51 years (*SD* = 0.52; range 4.6–6.2) in the TD group and 5.38 years (*SD* = 0.38; range 4.6–6.2) in the ASD group, *U* = 118.5, *p* = 0.46. There was no significant difference in language developmental age between the two groups.

Experimental Procedures

A narrative elicitation task was conducted by research assistants who were blind to the study design and hypotheses. The research assistants had been trained on the experimental procedures before the study. The instructions given to the research assistants were listed in **Table 1**. Six wordless pictures contained snapshots of a story about Tweety Bird and Sylvester were used. The story of Tweety Bird and Sylvester has been used in many prior studies

TABLE 1 | Instructions for experimenters in the narrative elicitation task.

Goal	Guideline/Example
1. Begin the story	"Let's begin now." "What's happing?" "One day…"
2. Draw children's attention	"Here is the next picture."
3. Encourage the children	a. Repeat children's speech "Yes, there is a bird." b. Use open questions "What is next?" "What is the end of the story?" c. Praise the children "Your story is lovely. Can you tell me more?"

to elicit speech and gestures. The story can be understood by both typical and atypical development across different cultures. Therefore, it is suitable for storytellers of different ages, different neurological conditions, and different language groups (McNeill, 1992, 2000). The story has a linear plot line about the two characters: Sylvester catches Tweety Bird, puts her in a sandwich, and tries to eat her.

During the narrative elicitation task, the research assistant presented the pictures to each child in a temporal order. Firstly, the research assistant invited the child to generate a story form some pictures. Then the child was given two minutes to look at the pictures. When telling the story, the researcher was also allowed to interact with the child. In this way, the narrative elicitation approximated a natural setting. The researcher encouraged the child to produce a story that was as long and complete as possible. However, the researcher was not allowed to produce any words or gestures related to the pictures. To reduce the demand on memory recall, the research assistant kept presenting the pictures while the child was narrating. In this narrative elicitation task, the child needed to extract a coherent narrative from the pictures and represent it linguistically (Botting, 2002). By generating a story from several wordless pictures, we minimized the demand for language comprehension and recall of the materials (Demir et al., 2010). The task was videotaped for later transcription and analyses.

Speech Transcription

Participants' spoken narratives were transcribed by trained coders who were native Cantonese speakers and blind to the hypotheses of the research study. All words and pauses were transcribed and further segmented into separate utterances, with each utterance containing a character and its corresponding action [e.g., "The cat catches the bird (*zi2 maau1 sik6 zi2 zoek3 zai2*)."]. Clauses with more than one character or action were broken into two or more utterances [e.g., "The cat eats the bread but the bird escapes (*zi2 maau1 ngaau5 go3 min6 baau1 daan6 hai6 bei2 zi2 zoek3 zai2 zau2 lat1 zo2*)]" was coded as two utterances as "The cat eats the bread (*zi2 maau1 ngaau5 go3 min6 baau1*)" and "The bird escapes (*zi2 zoek3 zai2 zau2 lat1 zo2*)". Utterances that did not contain information of the story were excluded from further analysis (e.g., "I have milk for breakfast."). All transcriptions were checked by a second trained coder who was also a native Cantonese speaker and blind to the hypotheses.

Gesture Coding

Identification of Gestures

All movements during narrations were coded by trained coders. The following movements were excluded: (1) hand movements that involved direct manipulation of an object (Goldin-Meadow et al., 1984); (2) motor stereotype and self-grooming movements (Silverman et al., 2017); (3) movements that did not contain information of the story (e.g., pointing to the fan on the wall).

Gesture Type

The present study followed a coding system initially described by McNeill (1992), who categorized co-speech gestures into four types: iconic, metaphoric, deictic, and beat. Iconic gestures resemble an aspect of the entity's shape or movement (e.g., both hands flapping to represent a bird flying). Metaphoric gestures convey abstract ideas or concepts (e.g., thumb and index finger moving toward each other while saying "The bread is a little hard."). Deictic or indexical gestures direct listeners' attention to the specified entities by pointing at them with an index finger (e.g., pointing to the sandwich while saying "The bird is inside."). Beat gestures are rhythmic hand movements that can segment and emphasize elements in speech (e.g., nodding while saying "Eat the bird."). Additionally, emblems, which can refer to culture-specific meanings as single words (e.g., horizontal shake of the head means "no") or phrases (e.g., shrugging the shoulders means "don't know") were also coded (de Marchena and Eigsti, 2010; Silverman et al., 2017).

Gesture Rate

Gesture rate was calculated as the number of gestures per utterance (total number of gestures divided by the total number of utterances).

Gesture Meaning and Gesture-Speech Relation

Each gesture was assigned a meaning based on its form and the co-occurring speech. The relationships between gesture meaning and co-occurring speech were categorized into three types depending on their semantic relationship (Özçalışkan and Goldin-Meadow, 2005; So et al., 2015; Wong and So, 2018). A reinforcing relation was coded when a gesture conveyed the same meaning as the co-occurring speech [e.g., shaking head when saying "The bird doesn't want to go out (*zi2 zoek3 zai2 m4 soeng2 ceot1 heoi3*)"]. A supplementary relationship was coded when a gesture added extra information. That is, the meaning of the gesture was not explicitly conveyed in the co-occurring speech [e.g., saying "The cat wants to eat the bird (*zi2 maau1 soeng2 sik6 zo2 zi2 zoek3 zai2*)" and producing a CATCH gesture]. A disambiguating relation was coded when a gesture clarified an underspecified referent [e.g., saying "The cat went there (*zi2 maau1 heoi3 zo2 go2 dou6*)" and pointing to the right]. The number of each type of gesture-speech relation was counted.

Reliability

To assess the inter-coder reliability, 20% of the cases were randomly selected and independently coded by a second trained coder. The inter-coder agreement was 0.96 ($N = 165$, Cohen's kappa = 0.96, $p < 0.001$) in an evaluation of gesture type and 0.86 ($N = 165$, Cohen's kappa = 0.86, $p < 0.001$) in an evaluation of gesture-speech relation.

Statistical Analyses

The Mann-Whitney test was used to examine differences in gesture rate, gesture type, and gesture-speech relation between the two groups.

RESULTS

Table 1 shows the proportion of different gesture types and gesture-speech relation. Around one-third of the gestures produced during the storytelling task were deictic gestures, while

TABLE 2 | Constitution of gestures produced by the ASD and TD group.

		Mean Proportion	
		ASD	TD
Gesture type	Deictic	33.3%	29.2%
	Iconic	48.3%	46.2%
	Metaphoric	1.1%	0%
	Beat	3.0%	2.7%
	Emblem	14.2%	21.9%
Gesture-speech relation	Reinforcing	73.8%	66.1%
	Supplementary	15.7%	22.6%
	Disambiguating	10.5%	11.3%
Deictic	Reinforcing	17.6%	17.3%
	Supplementary	5.2%	1.0%
	Disambiguating	10.5%	11.0%
Iconic	Reinforcing	38.6%	36.2%
	Supplementary	9.7%	10.0%
Emblem	Reinforcing	13.9%	10.6%
	Supplementary	0.4%	11.3%

iconic gestures accounted for about half of the total gestures in both groups. Most of the gestures (around 70%) represented a reinforcing meaning. Since the proportions of metaphoric gestures and beat gestures in gesture type were relatively small (less than 5%), they were excluded from the following analyses. **Figure 1** shows the average number of gestures by gesture type and gesture-speech relation in the two groups.

As shown in **Table 2**, both groups produced similar numbers of utterances during storytelling. However, the children with ASD produced significantly fewer gestures, resulting in a lower gesture rate compared to the TD group. In addition, the children with ASD produced fewer emblems and supplementary gestures, while the numbers of deictic gestures, iconic gestures, reinforcing gestures, and disambiguating gestures they produced were comparable to the TD group (**Table 3**).

We further analyzed the constitution of emblems by gesture-speech relation. Results showed that children with ASD tended to use emblems to reinforce accompanying speech (97.4%), while TD children did not show this tendency (51.5%). Besides, we analyzed the constitution of supplementary gestures by gesture type (deictic, iconic, and emblem). We found that in the TD group, half (50%) of the supplementary gestures were emblems, followed by 44.1% of iconic gestures. Deictic gestures only made up less than 5% (4.4%) of the supplementary gestures. In sharp contrast, only 2.4% of the supplementary gestures were emblems in the ASD group. Around two-thirds (61.9%) were iconic gestures and one-third (33.3%) were deictic gestures.

DISCUSSION

Results of the present study showed that the children with ASD had a lower gesture rate, which is consistent with the findings reported by So et al. (2015) and Silverman et al. (2017), whose participants were either school-age children or adolescents. In addition, echoing the findings of So et al. (2015) for school-age children, we found that the children with ASD produced fewer emblems than their TD peers, indicating that a delay in producing emblems exists in early and middle childhood. The children with ASD also had a delay in producing supplementary gestures, which was also reported by Morett et al. (2016) and So et al. (2015) in regard to autistic participants attending primary or middle school. These findings suggest that the impairment of gestural skills in individuals with ASD appears from preschool age and persists when they grow up.

By analyzing the constitution of emblems by gesture-speech relation, we found that most of the emblems produced by ASD had their meaning conveyed in the co-occurring speech, while TD produced half of the emblems without saying their meanings. Emblems, also known as conventional gestures, have culture-specific meanings and forms. However, children with ASD may

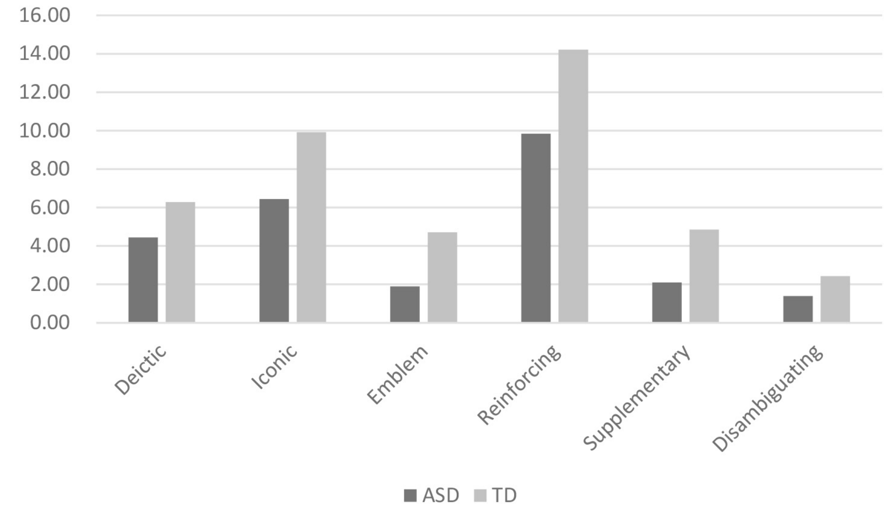


FIGURE 1 | Number of gestures by gesture type and gesture-speech relation.

TABLE 3 | Participants' characteristics, gestural skills, and comparison between the ASD and TD group.

	ASD (*n* = 20, 3 females)			TD (*n* = 14, 5 females)			Group comparison	
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**	***U***	***p*-value**
Chorological age	5.51	0.44	4.7–6.3	5.60	0.70	4.6–6.7	127	0.66
Language developmental age	5.38	0.38	4.6–6.2	5.51	0.52	4.6–6.2	118.5	0.46
Utterances[a]	30.85	10.46	11–59	35.50	12.26	16–63	106.5	0.25
Gestures[b]	13.35	8.07	2–30	21.50	10.11	10–42	71	0.02*
Gesture rate[c]	0.44	0.23	0.1–1.0	0.61	0.22	0.4–1.1	80	0.004**
Gesture type[d]								
Deictic	4.45	3.99	0-13	6.29	5.62	0–21	111	0.32
Iconic	6.45	4.26	0-14	9.93	6.15	2–23	94.5	0.11
Emblem	1.90	2.22	0-8	4.71	4.78	0–18	83	0.05*
Gesture-speech relation[e]								
Reinforcing	9.85	6.56	0–22	14.21	7.23	6–29	96	0.13
Supplementary	2.10	2.15	0–7	4.86	4.26	0–17	73.5	0.02*
Disambiguating	1.40	1.85	0–6	2.43	2.77	0–9	100.5	0.15
Deictic								
Reinforcing	2.35	2.35	0–7	3.71	3.71	0–12		
Supplementary	0.70	0.92	0–3	0.21	0.85	0–2		
Disambiguating	1.40	1.85	0–6	2.36	2.65	0–9		
Iconic								
Reinforcing	5.15	3.59	0–10	7.79	4.85	2–17		
Supplementary	1.30	1.75	0–5	2.14	1.99	0–6		
Emblem								
Reinforcing	1.85	2.23	0–8	2.29	2.09	0–7		
Supplementary	0.05	0.22	0–1	2.43	3.84	0–14		

*[a] Total number of utterances. [b] Total number of gestures. [c] Total number of gestures divided by the total number of utterances. [d] Number of each gesture type. [e] Number of each gesture-speech relation. *Significant at 0.05 level; **significant at 0.01 level.*

not realize that emblems could be produced and understood in a supplementary way. One possible explanation is that while children with ASD could learn some gestural skills from daily life as their TD peers (Wise and Sevcik, 2012), they are more likely to learn the gestures that are produced in a reinforcing way, in which the connection between the gesture and its corresponding meaning is explicit and clear (Knutsen et al., 2017). Therefore, they may have difficulty in learning emblems, which are more likely to be produced to supplement speech in daily life compared to other types of gestures (McNeill, 1992). Besides, children with ASD may be more likely to learn emblems that reinforce co-occurring speech, and produce them in the same way: reinforcing, rather than supplementary.

The delay in producing emblems may be a possible cause of impairment in producing supplementary gestures. Compared to other types of gestures, emblems can be used and understood without accompanying speech. These findings showed that compared to other types of gestures, TD children tended to produce emblems in a supplementary way, which is consistent with previous studies (McNeill, 1992, 2000). So et al. (2015) further pointed out that impairment in producing supplementary gestures may be due to the inability of individuals with ASD to integrate cross-modal semantic information. To produce gestures to supplement co-occurring speech, children have to coordinate information from both verbal language and hand movement, which may be more difficult for individuals with ASD than those with typical development. Notably, around one-third of supplementary gestures were deictic gestures in ASD. Producing deictic gestures to supplement speech (e.g., saying "eat" when pointing to the bread) is regarding as an early stage of development in both verbal language and gestures (Iverson and Goldin-Meadow, 2005; Özçalışkan et al., 2016). When children manage single words, they begin to use a gesture-plus-word combination (e.g., a verb + pointing) as two-word phrases, which is usually observed around 18 to 24 months in TD children (Iverson and Goldin-Meadow, 2005). Using deictic gestures in a supplementary way indicated that there may be some delay in gestural development in ASD.

However, unlike Camaioni et al. (2003) and Luyster et al. (2007), we did not find impairments in producing deictic and iconic gestures in the ASD group. Besides, de Marchena and Eigsti (2010) and de Marchena et al. (2019) reported no difference or marginally significant difference in gesture rate, which are not consistent with the results in this study. There are three possible reasons for these contradictory findings. One is about the task design. Some researchers have proposed that task design differences may result in variations across studies in language development, including narrative productions and gestural skills (Berman, 2004; Stirling et al., 2014). For example, de Marchena et al. (2019) found that participants in the ASD group used some types of gestures more often than those in the TD group in a collaborative referential communication task. These gestures were used to regulate turn-taking, which is not included in a storytelling task. Therefore, similar results may not be observed in the present study. Besides, some tasks may correlate with other social cognitive abilities. For example, asking children to retell a story to a stranger who had never read the story requires children's theory of mind understanding (Stirling et al., 2014). The last possible reason is age, Participants in de Marchena and Eigsti (2010) were adolescents and those in de Marchena et al. (2019) were adults, who may use gestures differently from preschool children. Therefore, it is critical to administer different tasks, as well as combine different findings, to obtain a better understanding of gestural skills in individuals with ASD. The second possible reason is the difference in the calculation of gesture rate. For example, de Marchena et al. (2019) defined gesture rate as the number of gestures per minute, while this study defined it as the number of gestures per utterance. In addition, the difference in the categories of gesture type is common. Apart from the gesture types used in this study, some researchers use categories including descriptive gestures, symbolic gestures, interactive gestures, and numerical gestures (Ingersoll, 2007; de Marchena et al., 2019). These differences in definition and characterization make it difficult to compare results across studies and to reach an agreement. Besides, the small sample size and inequality in sex ratio between the two groups may affect the results. Although we were not able to draw conclusions on the gestural impairment in ASD from this study, our findings show the differences in gestural use in TD and ASD. These findings could provide evidence for gestural training programs for children with ASD at an early age.

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 Institutional Review Board, the Chinese University of Hong Kong (Reference no. 14600817). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

YH: conceptualization, data curation, formal analysis, investigation, and writing – original draft. MW: data curation, project administration, and validation. W-YL: data curation. C-HC: data curation and validation. W-CS: conceptualization, methodology, funding acquisition, resources, supervision, and writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been fully supported by grants from the Chinese University of Hong Kong (Project no. CUHK3230211) and the Hong Kong Education Bureau (Project no. QEF6904338).

ACKNOWLEDGMENTS

We acknowledge the help of our research assistants, Phoebe Ng, Wai-Leung Wong, Chiu-Lok Ho, Kit-Ling Yeung, Cheuk-Chi Lee, Angel Pang, Christine Cheng, and Janis Man. Special thanks to all the children and their parents for their help and dedication to education.

REFERENCES

Berman, R. (2004). *Language Development Across Childhood and Adolescence*, Vol. 3. Amsterdam: John Benjamins Publishing.

Bochner, S., and Jones, J. (2003). *Child Language Development: Learning to Talk*, 2nd Edn. London: Whurr Publishers.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Lang. Teach. Ther.* 18, 1–21. doi: 10.1191/0265659002ct224oa

Camaioni, L., Perucchini, P., Muratori, F., Parrini, B., and Cesari, A. (2003). The communicative use of pointing in autism: developmental profile and factors related to change. *Eur. Psychiatry* 18, 6–12. doi: 10.1016/s0924-9338(02)00013-5

Charman, T., Drew, A., Baird, C., and Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur communicative development inventory (Infant Form). *J. Child Lang.* 30, 213–236. doi: 10.1017/s0305000902005482

Colgan, S. E., Lanter, E., McComish, C., Watson, L. R., Crais, E. R., and Baranek, G. T. (2006). Analysis of social interaction gestures in infants with autism. *Child Neuropsychol.* 12, 307–319. doi: 10.1080/09297040600701360

Colletta, J.-M., and Guidetti, M. (2012). *Gesture and Multimodal Development*. Amsterdam: John Benjamins Publishing Company.

de Marchena, A., and Eigsti, I.-M. (2010). Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. *Autism Res.* 3, 311–322. doi: 10.1002/aur.159

de Marchena, A., Kim, E. S., Bagdasarov, A., Parish-Morris, J., Maddox, B. B., Brodkin, E. S., et al. (2019). Atypicalities of gesture form and function in autistic adults. *J. Autism. Dev. Disord.* 49, 1438–1454. doi: 10.1007/s10803-018-3829-x

Demir, ÖE., Levine, S. C., and Goldin-Meadow, S. (2010). Narrative skill in children with early unilateral brain injury: a possible limit to functional plasticity. *Dev. Sci.* 13, 636–647. doi: 10.1111/j.1467-7687.2009.00920.x

Goldin-Meadow, S. (2006). Talking and thinking with our hands. *Curr. Dir. Psychol. Sci.* 15, 34–39. doi: 10.1111/j.0963-7214.2006.00402.x

Goldin-Meadow, S., Mylander, C., de Villiers, J., Bates, E., and Volterra, V. (1984). Gestural communication in deaf children: the effects and noneffects of parental input on early language development. *Monogr. Soc. Res. Child Dev.* 49, 143–151.

Hostetter, A. B. (2011). When do gestures communicate? a meta-analysis. *Psychol. Bull.* 137:297. doi: 10.1037/a0022128

Ingersoll, B. (2007). Teaching imitation to children with autism: a focus on social reciprocity. *J. Speech Lang. Pathol. Appl. Behav. Anal.* 2, 269–277. doi: 10.1037/h0100224

Iverson, J. M., and Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychol. Sci.* 16, 367–371. doi: 10.1111/j.0956-7976.2005.01542.x

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Lang. Cogn. Process.* 24, 145–167. doi: 10.1080/01690960802586188

Knutsen, J., Mandell, D. S., and Frye, D. (2017). Children with autism are impaired in the understanding of teaching. *Dev. Sci.* 20:e12368. doi: 10.1111/desc.12368

Lai, M.-C., Lombardo, M. V., and Baron-Cohen, S. (2014). Autism. *Lancet* 383, 896–910. doi: 10.1016/S0140-6736(13)61539-1

Luyster, R., Qiu, S., Lopez, K., and Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur–Bates Communicative Development Inventory. *J. Speech Lang. Hear Res.* 50, 667–681. doi: 10.1044/1092-4388(2007/047)

Mastrogiuseppe, M., Capirci, O., Cuva, S., and Venuti, P. (2015). Gestural communication in children with autism spectrum disorders during mother–child interaction. *Autism* 19, 469–481. doi: 10.1177/1362361314528390

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago press.

McNeill, D. (2000). "Catchments and contexts: non-modular factors in speech and gesture production," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press.), 312–328. doi: 10.1017/cbo9780511620850.019

Mitchell, S., Brian, J., Zwaigenbaum, L., Roberts, W., Szatmari, P., Smith, I., et al. (2006). Early language and communication development of infants later diagnosed with autism spectrum disorder. *J. Dev. Behav. Pediatr.* 27, S69–S78.

Morett, L. M. P. E., O'Hearn, K., Luna, B., and Ghuman, A. (2016). Altered Gesture and Speech Production in ASD Detract from In-Person Communicative Quality. *J. Autism Dev. Disord*. 46, 998–1012. doi: 10.1007/s10803-015-2645-9

Mundy, P., Sigman, M., and Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *J. Autism. Dev. Disord.* 20, 115–128. doi: 10.1007/BF02206861

Özçalışkan, Ş, Adamson, L. B., and Dimitrova, N. (2016). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism* 20, 754–763. doi: 10.1177/1362361315605921

Özçalışkan, Ş, and Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition* 96, B101–B113.

Schopler, E., Lansing, M., Reichler, R., and Marcus, L. (2005). *Examiner's Manual of Psychoeducational Profile*, Vol. 3. Austin, TX: Pro-ed Incorporation.

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., et al. (2015). Naturalistic developmental behavioral interventions: empirically validated treatments for autism spectrum disorder. *J. Autism. Dev. Disord.* 45, 2411–2428. doi: 10.1007/s10803-015-2407-8

Shek, D. T. L., and Yu, L. (2014). Construct validity of the Chinese version of the psycho-educational profile-(CPEP-3). *J. Autism. Dev. Disord.* 44, 2832–2843. doi: 10.1007/s10803-014-2143-5

Silverman, L. B., Eigsti, I.-M., and Bennetto, L. (2017). I tawt i taw a puddy tat: gestures in canary row narrations by high-functioning youth with autism spectrum disorder. *Autism Res.* 10, 1353–1363. doi: 10.1002/aur.1785

Smith, I. M. (1998). Gesture imitation in Autism I: nonsymbolic postures and sequences. *Cogn. Neuropsychol.* 15, 747–770. doi: 10.1080/026432998381087

So, W.-C., Wong, M. K.-Y., and Lam, K.-Y. (2016). Social and communication skills predict imitation abilities in children with Autism. *Front. Educ.* 1:3. doi: 10.3389/feduc.2016.00003

So, W.-C., Wong, M. K.-Y., Lam, W.-Y., Cheng, C.-H., Yang, J.-H., Huang, Y., et al. (2018). Robot-based intervention may reduce delay in the production of intransitive gestures in Chinese-speaking preschoolers with autism spectrum disorder. *Mol. Autism* 9:34. doi: 10.1186/s13229-018-0217-5

So, W.-C., Wong, M. K.-Y., Lui, M., and Yip, V. (2015). The development of co-speech gesture and its semantic integration with speech in 6- to 12-year-old children with autism spectrum disorders. *Autism* 19, 956–968. doi: 10.1177/1362361314556783

Stirling, L., Douglas, S., Leekam, S., and Carey, L. (2014). The use of narrative in studying communication in autism spectrum disorders. *Commun. Autism* 11, 169–216. doi: 10.1075/tilar.11.09sti

Stone, W. L., Ousley, O. Y., Yoder, P. J., Hogan, K. L., and Hepburn, S. L. (1997). Nonverbal communication in two-and three-year-old children with autism. *J. Autism. Dev. Disord.* 27, 677–696.

Tager-Flusberg, H., Paul, R., and Lord, C. (2005). Language and communication in autism. *Handb. Autism Pervasive Dev. Disord.* 1, 335–364.

Wetherby, A. M., Prizant, B. M., and Hutchinson, T. A. (1998). Communicative, social/affective, and symbolic profiles of young children with autism and pervasive developmental disorders. *Am. J. Speech Lang. Pathol.* 7, 79–91. doi: 10.1044/1058-0360.0702.79

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., and Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *J. Autism. Dev. Disord.* 34, 473–493. doi: 10.1007/s10803-004-2544-y

Wise, J. C., and Sevcik, R. A. (2012). "Language development," in *Encyclopedia of Human Behavior*, ed. V. S. Ramachandran (San Diego: Academic Press), 511–516.

Wong, M. K.-Y., and So, W.-C. (2018). Absence of delay in spontaneous use of gestures in spoken narratives among children with Autism Spectrum Disorders. *Res. Dev. Disabil.* 72, 128–139. doi: 10.1016/j.ridd.2017.11.004

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

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 17 September 2020
doi: 10.3389/fpsyg.2020.571109

Does Gesture Lighten the Load? The Case of Verbal Analogies

Acacia L. Overoye[1]* *and Margaret Wilson*[2]

[1]Behavioral Science Department, Utah Valley University, Orem, UT, United States, [2]Psychology Department, University of California Santa Cruz, Santa Cruz, CA, United States

Gesturing has been shown to relay benefits to speakers and listeners alike. Speakers, for instance, may be able to reduce their working memory load through gesture. Studies with children and adults have demonstrated that gesturing while describing how to solve a problem can help to save cognitive resources related to that explanation, allowing them to be allocated to a secondary task. The majority of research in this area focuses on procedural mathematical problem solving; however, the present study examines how gesture interacts with working memory load during a verbal reasoning task: verbal analogies. Unlike previous findings which report improved performance on secondary tasks while gesturing during a primary task, our results show that participants showed better performance in a secondary memory task when being prohibited from gesturing during their explanation of verbal analogies compared to being allowed to gesture. These results suggest that the relationship between gesture and working memory may be more nuanced, with the type of task and gestures produced influencing how gestures interact with working memory load.

Keywords: gesture, working memory, cognitive load, offloading, problem solving

OPEN ACCESS

Edited by:
Naomi Sweller,
Macquarie University, Australia

Reviewed by:
Tilbe Goksun,
Koç University, Turkey
Myrto Mavilidi,
University of Wollongong, Australia

****Correspondence:***
Acacia L. Overoye
acacia.overoye@uvu.edu

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

Received: *09 June 2020*
Accepted: *24 August 2020*
Published: *17 September 2020*

Citation:
Overoye AL and Wilson M (2020)
Does Gesture Lighten the Load? The
Case of Verbal Analogies.
Front. Psychol. 11:571109.
doi: 10.3389/fpsyg.2020.571109

INTRODUCTION

People spontaneously produce hand movements, gestures, alongside speech. The use of gesture is cross-cultural and individuals from different backgrounds produce gestures tied to their cultural and linguistic heritage (Kendon, 1995; Kita, 2009). The gestures speakers produce are not mere hand-waving but confer benefits to listeners and speakers alike (Novack and Goldin-Meadow, 2015; Dargue et al., 2019). Gesturing while speaking has been found to facilitate problem solving (Cook and Tanenhaus, 2009; Beilock and Goldin-Meadow, 2010; Chu and Kita, 2011; Eielts et al., 2018), learning and memory (Stevanoni and Salmon, 2005; Broaders et al., 2007; Goldin-Meadow et al., 2009; Stieff et al., 2016), and speech production and organization (Graham and Heywood, 1975; Rauscher et al., 1996; Morsella and Krauss, 2004; Hostetter et al., 2007; Jenkins et al., 2017). Gesture has also been shown to improve comprehension, and this enhancement extends across age groups (Dargue et al., 2019). Some have suggested that the beneficial effects of gesture on problem solving and learning are related to how gesture can assist in managing working memory load (Goldin-Meadow and Wagner, 2005; Goldin-Meadow, 2011).

Individual differences in working memory can influence the relationship between gesture use and comprehension. Individuals with lower visuospatial and verbal working memory capacity have been found to produce co-speech gestures more frequently (Chu et al., 2014; Gillespie et al., 2014; Pouw et al., 2016). On the side of comprehension, individuals appear to be more sensitive to information conveyed in gesture when they have higher visuospatial working memory capacity

(Wu and Coulson, 2014a,b; Özer and Göksun, 2019). Not only is the extent to which an individual produces gesture and their sensitivity to gesture influenced by their working memory capacity, research has also shown that the production of gesture can change how an individual uses working memory.

Goldin-Meadow et al. (2001) studied the relationship between gesture production and working memory load in a dual-task paradigm with both children and adults. Adults were given a primary task of solving and explaining math problems [factoring polynomials such as $x^2 + 4x + 4 = () ()$], while completing a secondary memory task of remembering letters. In each trial of Goldin-Meadow et al. study, participants first solved a factoring problem and were then presented with letters to remember. Participants then explained how they solved the factoring problem and were either permitted to move their hands or required to keep them still while speaking. Finally, participants recalled the set of letters. Results showed that when the participants gestured during their explanation, they subsequently recalled the letters more accurately compared to explanations where they did not gesture. Goldin-Meadow et al. (2001) explained that gesturing reduced working memory load during explanations, resulting in a greater allocation of cognitive resources to the maintenance of letters in working memory for the secondary task. Further research has demonstrated that co-speech gestures manage working memory load more effectively than meaningless hand-waving (Cook et al., 2012), in visuospatial working memory tasks (Wagner et al., 2004), and when gestures refer to problems that are not in the present environment (Ping and Goldin-Meadow, 2010). Additionally, the production of co-speech gestures during explanations appears to be especially effective at reducing working memory load for individuals with low working memory capacity (Marstaller and Burianová, 2013).

There are several theoretical explanations for how gesture reduces working memory load. Producing gestures may help speakers to simulate visuospatial and motoric representations more easily, thereby freeing up additional resources that would have otherwise been necessary for creating simulations (Hostetter and Alibali, 2008; Ping and Goldin-Meadow, 2010; Risko and Gilbert, 2016). Alternatively, the production of gesture may provide speakers with externalized frameworks for problem solving and assist in reducing load by chunking mental work into manageable units (Kita, 2000). Gesturing may also help a speaker shift load from verbal working memory to other, visuospatial, or motoric representations (Paas and Sweller, 2012).

While these studies suggest that gesture alleviates working memory load during speech, much of the dual-task research on gesture and working memory load has focused on mathematical problem solving (Goldin-Meadow et al., 2001; Wagner et al., 2004; Cook et al., 2012; Marstaller and Burianová, 2013). There are several reasons to question whether the relationship between gesture and working memory load found in mathematical problem solving will generalize to other types of problem solving. First, finding and explaining the solution to math problems is often a procedural process. Consider factoring problems: although the numbers in the polynomial vary, the steps a problem solver goes through to factor the polynomial are consistent across problems. This property of factoring problems makes them particularly well suited to benefit from gestures. The gestures produced during the explanation of a factoring problem serve as a repeated structural hangar for speech, saving cognitive resources for a secondary task. In other types of problem solving, gestures may be germane to the specific contents of each problem. Instead of reinforcing a repeated procedure, gestures in other types of problem solving may illustrate unique relationships or surface level details that change across problems.

A second point is that different tasks may elicit different types of gestures, and these may interact with working memory in unique ways. The gestures produced by speakers while explaining factoring problems are primarily deictic (see Wagner et al., 2004; for examples) and the working memory load reduction observed for the speaker may be a result of linking information in speech to representations in the present environment (Ping and Goldin-Meadow, 2010). Although this point was addressed in Ping and Goldin-Meadow (2010) where it was found that iconic gestures produced during explanations of conservation problems about non-present objects can reduce working memory load for children, it is an open question whether the same holds true for adults and for other tasks which elicit different types of gesture. As other work has found that the gestures speakers produce during other types of reasoning tasks are associated with individual differences in working memory (Chu et al., 2014), it seems important to investigate whether the various gesture elicitation tasks result in different effects of gesture production on working memory load.

Finally, the math problems used in previous research rely on numbers – symbols that do not necessarily have strong visuospatial features that could potentially mislead problem solvers. When explaining a math problem, gestures can index numbers and how they relate to each other with less of a chance to introduce irrelevant information into solving the problem. Previous research has found that for certain problem-solving tasks (analogical problem solving, Tower of Hanoi), gestures can interfere with coming to a correct solution because they introduce irrelevant information (Trofatter et al., 2015; Hostetter et al., 2016).

In the present study, we adopt the dual-task paradigm recruited in previous research (e.g., Goldin-Meadow et al., 2001; Wagner et al., 2004; Cook et al., 2012) and investigate whether a different primary task, verbal analogies, influences the relationship between gesture and working memory load. One reason for choosing analogies is that simply, verbal analogies are a different type of task than the spatial and mathematical problem-solving tasks that are often used in this type of work. Gestures can represent analogical relationships (Cooperrider and Goldin-Meadow, 2017) and have the potential to encourage a variety of different iconic, deictic, and beat gestures. Further, there is no set procedural formula to solve an analogy and they require the problem solver to consider non-abstract information.

EXPERIMENT 1A

Experiment 1a used a dual-task paradigm to investigate the influence of gesture production during a primary verbal analogy task on a secondary memory task. Similar to previous gesture and working memory work using this paradigm (e.g., Goldin-Meadow et al., 2001; Wagner et al., 2004; Cook et al., 2012), performance on the secondary memory task served as measure of working memory load. Better performance on the secondary memory task when producing gestures as opposed to being prohibited from gesture during the verbal analogy task would indicate that gestures assist in alleviating working memory load. Alternatively, improved performance on the secondary memory task while being prohibited from gesturing during the verbal analogy task would indicate that producing gestures does not assist in reducing working memory load for the analogy task.

Method

Participants

Forty-four undergraduates from the University of California, Santa Cruz (UCSC) participated for partial course credit. Participants were recruited from psychology courses at UCSC through the Sona Systems subject pool and were required to be native speakers of English to be eligible for the study. Four participants were removed from the analysis because they did not write down responses to the task or did not complete the experiment. The study was reviewed and approved by the UCSC IRB. The participants provided their written informed consent to participate in this study.

Design

The experiment used a within-subjects design with gesture instruction as the independent variable (gesture encouraged vs. gesture prohibited) and performance on the secondary memory task as the dependent variable.

Materials

Forty verbal analogies were selected for use in the study. All analogies followed the form, "A is to B as C is to …" such as "Hat is to head as roof is to …" Analogies were written such that they relied on many different types of relationships between analogous items such as color, shape, movement, and spatial relationships. Analogies were chosen after a pilot phase where 20 participants answered 50 analogies and were scored for accuracy. The most challenging 10 analogies were removed and the remaining 40 were solved by participants with an accuracy of 65% (*SD* = 22%). These 40 analogies were divided into two lists of 20. A list of all analogies used in the experiment is available in the **Supplementary Material**.

Procedure

Participants were presented with the experiment on Psyscope – a graphical user interface (GUI) program used to develop psychology experiments (Cohen, 1993). Before beginning the experiment, participants were provided with both verbal and written instructions that indicated they would be completing a verbal analogy and memory task. Participants were shown an example analogy with its solution and completed one example trial of the experiment. Participants were also informed that they would receive instructions on how to position their hands during different phases in the experiment.

After receiving the instructions, participants were presented with the first of two counterbalanced blocks. In each block, participants were presented with an analogy and were given unlimited time to solve it. Once participants solved the analogy, they pressed a button and were presented with a list of six pseudorandom numbers for the memory task. After viewing the numbers for 5 s, the original analogy returned to the screen and participants explained how they arrived at their answer. After finishing their explanation, participants were asked to recall the six numbers by writing them on a worksheet. The participants completed this process for 20 analogies in each of the two blocks.

Before each block participants were given instructions on how to position their hands. In previous research, multiple instructions have been used to encourage and discourage gesture use such as, permitting and not permitting movement (Goldin-Meadow et al., 2001) or directly asking participants to gesture (Cook et al., 2012) without a change in results. We chose to explicitly instruct participants to gesture to increase the likelihood of gesture production in the gesture encouraged condition. For the gesture prohibited instructions, participants were instructed to keep their hands flat and still on the table in front of them. If the experimenter noticed a participant not following the gesture instructions, they gently reminded the participants to keep their hands still or gesture as needed. Instructions and blocks were counterbalanced such that the order and pairing of gesture instruction and analogy list occurred in all possible combinations across participants. Participants completed the entirety of the experiment in under an hour.

Results and Discussion

A paired-samples *t*-test was conducted to compare performance on the secondary memory task in the gesture encouraged and gesture prohibited conditions. Results showed that participants remembered more digits when being prohibited from gesturing ($M = 0.42$, $SE = 0.3$) than being encouraged to gesture ($M = 0.39$, $SE = 0.03$), $t(39) = 2.15$, $p = 0.04$, $d = 0.34$. These results demonstrate a reversal in previous findings (e.g., Goldin-Meadow et al., 2001; Wagner et al., 2004; Cook et al., 2012) – producing gestures resulted in worse performance on the secondary memory task than being prohibited from gesturing. This indicates that gestures produced while explaining verbal analogies may not free up resources in working memory.

We conducted several follow-up analyses to assess the influence of order of gesture instructions and item effects. First, we evaluated whether the order of the two instruction conditions (gesture encouraged and gesture prohibited) influenced recall. A repeated measures ANOVA with gesture instruction as a within-subjects factor and order of conditions as a between-subjects factor revealed a main effect of gesture instruction [$F(1,38) = 4.93$, $p = 0.03$, $\eta^2 = 0.12$], but no main effect of order [$F(1,38) = 0.51$, $p = 0.48$, $\eta^2 = 0.01$] or interaction between order and gesture instruction [$F(1,38) = 2.36$, $p = 0.13$, $\eta^2 = 0.06$]. This indicates that irrespective the order of instructions, being instructed to gesture resulted in lower performance on digit recall compared to instructions to not gesture.

A univariate general linear model (GLM) analysis was conducted to examine whether the influence of gesture instruction on recall persists when controlling for variability from the analogy items and differences across participants. The model included gesture instruction as a fixed factor, and analogy item (nested within gesture instruction) and participant as random factors. The GLM analysis revealed significant main effects of gesture instruction (F = 5.09, p = 0.027) and participant (F = 10.01, p < 0.01), but not analogy (F = 1.28, p = 0.054). A summary of the analysis is available in **Table 1**. These results show that participants performed better on the memory task when prohibited from gesturing rather than being encouraged to gesture, even when controlling for item and participant variability.

Perhaps some analogies in this study were better suited to gesturing than others, and that gesture only reduces working memory load for concepts that are readily gestured about. Although people produce gestures while speaking about all kinds of information, previous research has shown that gestures are produced more frequently and consistently for speech that has content related to visuospatial information (Krauss, 1998; Alibali et al., 2001). Theories, such as the gesture as simulated action framework (Hostetter and Alibali, 2008, 2019), further argue that gestures are produced in part as result of visuospatial simulations. It is possible then that only analogies that depict visuospatial relationships would show a benefit of gesturing during explanations for working memory. To examine this possibility, we divided the analogies into two categories: analogies which focused on spatial relationships and shapes (e.g., belt is to waist as equator is to? and kite is to diamond as egg is to?) and analogies unrelated to special relationships such as those about color (e.g., apple is to banana as red is to?) and conducted a repeated measures ANOVA with gesture instructions (gesture encouraged and gesture prohibited) and analogy type (spatial and other) as within-subjects variables. The analysis revealed a significant main effect of gesture instruction [F(1, 38) = 5.14, p = 0.029, η^2 = 0.12], but no main effect of analogy type [F(1, 38) = 0.003, p = 0.95, η^2 < 0.00] or interaction between gesture instruction and analogy type [F(1, 38) = 2.29, p = 0.014, η^2 = 0.06]. These results indicated that irrespective of the type of analogy, participants performed better on the secondary memory task when being prohibited from gesture rather than being encouraged to gesture.

EXPERIMENT 1B

Experiment 1a showed that unlike previous research where gesture production has been shown to help manage working memory load (e.g., Goldin-Meadow et al., 2001), being instructed to gesture during the explanation of verbal analogies did not help reduce working memory load and instead led to worse performance when compared to being prohibited from gesturing. Given this surprising finding, Experiment 1b was conducted to replicate the main finding of Experiment 1a. Additionally, Experiment 1b employed a few small changes to eliminate potential participant fatigue, distraction, and more closely align with the methods used in previous research.

TABLE 1 | Univariate general linear model (GLM) analysis between gesture instruction, analogy, and participant on recall.

	df	*F*	*p*
Gesture instruction	1	5.09	0.027
Participant	38	10.01	<0.01
Analogy	77	1.28	0.054

The model considered gesture instruction as a fixed factor and participant and analogy as random factors.

Method

Participants

Twenty-one undergraduates from UCSC participated for partial course credit. Participants were recruited from psychology courses at UCSC through the Sona Systems subject pool and were required to be native speakers of English to be eligible for the study. The study was reviewed and approved by the UCSC and participants provided their written informed consent to participate in this study.

Materials

Materials consisted of a subset of analogies from Experiment 1a. A full list is available in the **Supplementary Material**.

Procedure

The procedure was the same as Experiment 1a with few minor changes. Participants were fewer verbal analogies (15 in each condition) to eliminate potential effects of fatigue. Additionally, the secondary memory task was changed to pseudorandom consonants (e.g., v, r, k, p, q, and d) instead of numbers to match the task used by Goldin-Meadow et al. (2001). Finally, participants entered their responses to the secondary task with the keyboard instead of on paper to reduce potential costs of switching between using the computer and paper.

Results and Discussion

Experiment 1b showed the same pattern of results as Experiment 1a, with participants recalling more consonants when instructed to prohibit gesture (M = 0.38, SE = 0.4) rather than produce gestures (M = 0.33, SE = 0.04), t(20) = 2.16, p = 0.04, d = 0.47. A repeated measures ANOVA with instruction order as a between-subjects factor found a main effect of gesture instruction [F(1,19) = 4.52, p = 0.047, η^2 = 0.192] but no effect of order [F(1,19) = 0.49, p = 0.49, η^2 = 0.03] and no interaction between order and instruction [F(1,19) = 1.40, p = 0.25, η^2 = 0.07]. A summary of the results of both experiments can be seen in **Table 2**.

TABLE 2 | Mean proportion recalled for Experiments 1a and 1b.

	Gesture instruction	
Experiments	**Gesture encouraged**	**Gesture prohibited**
Experiment 1a	0.39 (0.03)	0.42 (0.03)
Experiment 1b	0.33 (0.04)	0.38 (0.04)

Standard error in parentheses.

GENERAL DISCUSSION

Gesturing during explanations has previously been shown to alleviate working memory load (Goldin-Meadow et al., 2001; Wagner et al., 2004; Cook et al., 2012). These studies have used dual-task paradigms to show that when individuals gesture during an explanation task their performance on a secondary memory task is enhanced compared to explaining without using gestures. These findings demonstrate that a speaker's own gestures can influence their working memory and allow speakers to allocate cognitive resources that would have otherwise been used in an explanation to a secondary task.

The aim of the present research was to build on the previous research of Goldin-Meadow et al. (2001), Wagner et al. (2004), Cook et al. (2012), and others to explore whether gesture during a novel verbal analogy task would have similar effects on working memory load. Unlike previous research, our studies showed that gesturing during the explanation of verbal analogies did not lighten working memory load. Instead, being instructed to gesture led to worse performance on a secondary memory task when compared to being prohibited from gesturing. These results suggested that although producing gestures may help manage working memory load in some contexts, it may create additional load in working memory in other contexts.

There are several possible explanations for why gesture did not reduce working memory load during the explanation of analogies. For one, explaining and gesturing about verbal analogies may have led participants to use cognitive resources to build visuospatial representations of the content of the analogies. According to the gesture as simulated action framework (Hostetter and Alibali, 2008, 2019), gestures emerge from embodied visuospatial and motor representations used in speaking and thinking. For an example, consider solving the analogy, "belt is to waist, as equator is to …" In the explanation, a participant could explain that belts go around the waist, as an equator goes around the globe and produce a gesture of one hand circling around another or around the participant's body. The "going around" gesture is an emergent action of the motor representation of the visuospatial concept "going around" that is needed to solve the analogy problem. Creating and maintaining this representation in mind in service of producing a gesture could add more load than being prohibited from gesture and not needing to construct such vivid visuospatial representations.

Another possibility is that the different ways participants recruit gestures in their explanations could have varying effects on working memory. In explaining an analogy, gestures could be used to highlight different relationships that are key for solving the analogy that differ across problems, index words on the screen, or provide rhythm to the explanatory speech. These usages of gesture may inconsistently interact with working memory load with some gestures being more effective than others and freeing up cognitive resources. While we do not have video data to explore these possibilities in the present study, investigating the association between the gesture strategy used by participants and extent of working memory load could help clarify these relationships in future research.

The difficulty and unfamiliarity of analogies may have also influenced the results of our studies. Verbal analogies are not typically taught in schooling and participants may have had little experience solving and providing explanations for analogies. Our analogies were more challenging for participants than factoring problems used in previous research and participants did not have a means to "check their answers" to see if they reached the correct solution. The potential difficulty and unfamiliarity with the task could have created additional load for participants and influenced results. However, since the analogies and gesture instruction condition were counterbalanced across participants in present study, we believe that factors such as difficulty alone cannot explain the observed difference between the gesture encouraged and gesture prohibited conditions.

There are several limitations to consider when interpreting the results of this research. First, our manipulation consisted of instructing participants to produce or prohibit gestures. Although this gave us a clear comparison between two conditions, it also removed the possibility of determining a baseline rate of gesture for each of our participants. It is possible that individuals who gesture more in their day-to-day lives may differentially benefit from gesture use than individuals who gesture infrequently. Encouraging our participants to gesture may have also been more distracting from the secondary memory task than prohibiting them from gesturing. Differences in recall could have been due to the added difficulty of remembering to produce gestures which may have been greater than the difficulty of inhibiting gesture. Additionally, the absence of video data limits our ability to compare our results with previous studies and analyze how specific gestures and the consistency of the gestures produced may have influenced offloading. Follow-up research could elaborate on the verbal analogies themselves by developing analogies that have been evaluated for reliability and investigating how gesture use influences the accuracy of solving the analogies and memory for the solutions. Future research can also examine the background of participants and consider whether age or other demographic factors influence the results.

These findings highlight the need for a nuanced approach to studying the relationship between gesture and other cognitive processes. While the gestures produced in the explanation of mathematical and other types of problems may save cognitive resources, those produced during the explanation of analogies may have the opposite effect. Similarly, although gesture can aid in problem solving in some domains (such as mental rotation), it can bias a problem solver and lead to less efficient or incorrect problem solving in other scenarios (Alibali et al., 2011; Göksun et al., 2013; Hostetter et al., 2016). Both the gestures and the context in which they are produced may influence the extent to which gesture is beneficial to a speaker.

The results of this initial work on verbal analogies and gesture indicates that the content gestures refer to may matter in its relationship to working memory load. Gestures may not interact with all spoken content equally, but instead may adapt to the constraints of a situation. This work adds to a growing literature that demonstrates context, individual differences, and type of gesture influence how gesture production interacts with cognition.

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/bja5c/?view_only=0bdd50a2aa664e45ada5b4b7a0b807bf – Open Science Framework.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of California, Santa Cruz Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AO and MW contributed to conception and design of the study. AO organized and analyzed the data and wrote the initial draft of the manuscript. MW provided feedback on the manuscript and approved of a final version that was completed by AO. All authors contributed to the article and approved the submitted version.

FUNDING

Publication fees paid by Behavioral Science department at Utah Valley University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.571109/full#supplementary-material

REFERENCES

Alibali, M. W., Heath, D. C., and Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *J. Mem. Lang.* 44, 169–188. doi: 10.1006/jmla.2000.2752

Alibali, M. W., Spencer, R. C., Knox, L., and Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychol. Sci.* 22, 1138–1144. doi: 10.1177/0956797611417722

Beilock, S. L., and Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychol. Sci.* 21, 1605–1610. doi: 10.1177/0956797610385353

Broaders, S. C., Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *J. Exp. Psychol.* 136, 539–550. doi: 10.1037/0096-3445.136.4.539

Chu, M., and Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *J. Exp. Psychol.* 140, 102–116. doi: 10.1037/a0021790

Chu, M., Meyer, A., Foulkes, L., and Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *J. Exp. Psychol. Gen.* 143, 694–709. doi: 10.1037/a0033861

Cohen, J. D. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behav. Res. Methods* 25, 257–271. doi: 10.3758/BF03204507

Cook, S. W., and Tanenhaus, M. K. (2009). Embodied communication: speakers' gestures affect listeners' actions. *Cognition* 113, 98–104. doi: 10.1016/j.cognition.2009.06.006

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Lang. Cogn. Process.* 27, 594–610. doi: 10.1080/01690965.2011.567074

Cooperrider, K., and Goldin-Meadow, S. (2017). When gesture becomes analogy. *Top. Cogn. Sci.* 9, 719–737. doi: 10.1111/tops.12276

Dargue, N., Sweller, N., and Jones, M. P. (2019). When our hands help us understand: a meta-analysis into the effects of gesture on comprehension. *Psychol. Bull.* 145, 765–784. doi: 10.1037/bul0000202

Eielts, C., Pouw, W., Ouwehand, K., Van Gog, T., Zwaan, R. A., and Paas, F. (2018). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychol. Res.* 84, 502–513. doi: 10.1007/s00426-018-1065-9

Gillespie, M., James, A. N., Federmeier, K. D., and Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: evidence from individual differences. *Cognition* 132, 174–180. doi: 10.1016/j.cognition.2014.03.012

Göksun, T., Goldin-Meadow, S., Newcombe, N., and Shipley, T. (2013). Individual differences in mental rotation: what does gesture tell us? *Cogn. Process.* 14, 153–162. doi: 10.1007/s10339-013-0549-1

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdiscip. Rev. Cogn. Sci.* 2, 595–607. doi: 10.1002/wcs.132

Goldin-Meadow, S., Cook, S. W., and Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychol. Sci.* 20, 267–272. doi: 10.1111/j.1467-9280.2009.02297.x

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395

Goldin-Meadow, S., and Wagner, S. M. (2005). How our hands help us learn. *Trends Cogn. Sci.* 9, 234–241. doi: 10.1016/j.tics.2005.03.006

Graham, J. A., and Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *Eur. J. Soc. Psychol.* 5, 189–195. doi: 10.1002/ejsp.2420050204

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495

Hostetter, A. B., and Alibali, M. W. (2019). Gesture as simulated action: revisiting the framework. *Psychon. Bull. Rev.* 26, 721–752. doi: 10.3758/s13423-018-1548-0

Hostetter, A. B., Alibali, M. W., and Kita, S. (2007). I see it in my hands' eye: representational gestures reflect conceptual demands. *Lang. Cogn. Process.* 22, 313–336. doi: 10.1080/01690960600632812

Hostetter, A. B., Wieth, M., Foster, K., Moreno, K., and Washington, J. (2016). "Effects of gesture on analogical problem solving: when the hands lead you astray" in *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. eds. A. Papafragou, D. Grodner, D. Mirman and J. C. Trueswell (Cognitive Science Society), 1685–1690.

Jenkins, T., Coppola, M., and Coelho, C. (2017). Effects of gesture restriction on quality of narrative production. *Gesture* 16, 416–431. doi: 10.1075/gest.00003.jen

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversation. *J. Pragmat.* 23, 247–279. doi: 10.1016/0378-2166(94)00037-F

Kita, S. (2000). How representational gestures help speaking. *Lang. Gesture* 1, 162–185. doi: 10.1017/CBO9780511620850.011

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Lang. Cogn. Process.* 24, 145–167. doi: 10.1080/01690960802586188

Krauss, R. M. (1998). Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* 7, 54–60. doi: 10.1111/1467-8721

Morsella, E., and Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *Am. J. Psychol.* 117, 411–424. doi: 10.2307/4149008

Marstaller, L., and Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychon. Bull. Rev.* 20, 496–500. doi: 10.3758/s13423-012-0365-0

Novack, M., and Goldin-Meadow, S. (2015). Learning from gesture: how our hands change our minds. *Educ. Psychol. Rev.* 27, 405–412. doi: 10.1007/s10648-015-9325-3

Özer, D., and Göksun, T. (2019). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Lang. Cogn. Neurosci.* 1–19. doi: 10.1080/23273798.2019.1703016

Paas, F., and Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educ. Psychol. Rev.* 24, 27–45. doi: 10.1007/s10648-011-9179-2

Ping, R., and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cogn. Sci.* 34, 602–619. doi: 10.1111/j.1551-6709.2010.01102.x

Pouw, W. T., Mavilidi, M. F., Van Gog, T., and Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cogn. Process.* 17, 269–277. doi: 10.1007/s10339-016-0757-6

Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). Gesture, speech, and lexical access: the role of lexical movements in speech production. *Psychol. Sci.* 7, 226–231. doi: 10.1111/j.1467-9280.1996.tb00364.x

Risko, E. F., and Gilbert, S. J. (2016). Cognitive offloading. *Trends Cogn. Sci.* 20, 676–688. doi: 10.1016/j.tics.2016.07.002

Stevanoni, E., and Salmon, K. (2005). Giving memory a hand: instructing children to gesture enhances their event recall. *J. Nonverbal Behav.* 29, 217–233. doi: 10.1007/s10919-005-7721-y

Stieff, M., Lira, M. E., and Scopelitis, S. A. (2016). Gesture supports spatial thinking in STEM. *Cogn. Instr.* 34, 80–99. doi: 10.1080/07370008.2016.1145122

Trofatter, C., Kontra, C., Beilock, S., and Goldin-Meadow, S. (2015). Gesturing has a larger impact on problem-solving than action, even when action is accompanied by words. *Lang. Cogn. Neurosci.* 30, 251–260. doi: 10.1080/23273798.2014.905692

Wagner, S. M., Nusbaum, H., and Goldin-Meadow, S. (2004). Probing the mental representation of gesture: is handwaving spatial? *J. Mem. Lang.* 50, 395–407. doi: 10.1016/j.jml.2004.01.002

Wu, Y. C., and Coulson, S. (2014a). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychol.* 153, 39–50. doi: 10.1016/j.actpsy.2014.09.002

Wu, Y. C., and Coulson, S. (2014b). A psychometric measure of working memory capacity for configured body movement. *PLoS One* 9:e84834. doi: 10.1371/journal.pone.0084834

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 September 2020
doi: 10.3389/fpsyg.2020.574418

Emblem Gestures Improve Perception and Evaluation of Non-native Speech

*Kiana Billot-Vasquez[1,2], Zhongwen Lian[2,3], Yukari Hirata[2,3,4] and Spencer D. Kelly[1,2,3]**

[1] Department of Psychological and Brain Sciences, Colgate University, Hamilton, NY, United States, [2] Center for Language and Brain, Hamilton, NY, United States, [3] Linguistics Program, Colgate University, Hamilton, NY, United States, [4] Department of East Asian Languages, Colgate University, Hamilton, NY, United States

Traditionally, much of the attention on the communicative effects of non-native accent has focused on the accent *itself* rather than how it functions within a more natural context. The present study explores how the bodily context of co-speech emblematic gestures affects perceptual and social evaluation of non-native accent. In two experiments in two different languages, Mandarin and Japanese, we filmed learners performing a short utterance in three different within-subjects conditions: speech alone, culturally familiar gesture, and culturally unfamiliar gesture. Native Mandarin participants watched videos of foreign-accented Mandarin speakers (Experiment 1), and native Japanese participants watched videos of foreign-accented Japanese speakers (Experiment 2). Following each video, native language participants were asked a set of questions targeting speech perception and social impressions of the learners. Results from both experiments demonstrate that familiar—and occasionally unfamiliar—emblems facilitated speech perception and enhanced social evaluations compared to the speech alone baseline. The variability in our findings suggests that gesture may serve varied functions in the perception and evaluation of non-native accent.

Keywords: speech processing, non-native accent, hand gesture, multimodal, second language, cross-cultural communication

OPEN ACCESS

Edited by:
Kazuki Sekine,
Waseda University, Japan

Reviewed by:
Laura M. Morett,
University of Alabama, United States
Andrej A. Kibrik,
Institute of Linguistics (RAS), Russia

****Correspondence:***
Spencer D. Kelly
skelly@colgate.edu

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

Received: *19 June 2020*
Accepted: *19 August 2020*
Published: *22 September 2020*

Citation:
Billot-Vasquez K, Lian Z, Hirata Y and Kelly SD (2020) Emblem Gestures Improve Perception and Evaluation of Non-native Speech. Front. Psychol. 11:574418. doi: 10.3389/fpsyg.2020.574418

INTRODUCTION

More than half of the world's population is bilingual, a pattern that has only accelerated since the turn of the millennium (Grosjean, 2010). Studies focused on the treatment and perception of non-native accented speech have shown that it is consistently discriminated against, negatively affecting measures related to likeability, sociability, and intelligence (Bradac, 1990; Lindemann, 2003). In an effort to understand accented speech within a natural communicative context, the present study explores how non-native accents are perceived and evaluated in the presence of co-speech emblematic gestures. Building on research demonstrating that gesture's semantic relationship with speech can powerfully affect language processing, comprehension and learning (Church et al., 2017), the present study asks how a gesture's *cultural* relationship to speech influences cross-cultural perceptions and impressions of accented speech and speakers.

The Stigma of Accent

Many people learn their non-native language later in life—through formal education or pressures from commerce—so it is commonplace to speak a second language with a non-native accent (Johnson and Newport, 1989; Cheng, 1999). In general, a non-native accent, a term interchangeable with foreign accent, has been defined as "speech that systematically diverges from native speech due to interference from the phonological and acoustic-phonetic characteristics of a talker's native language" (Atagi and Bent, 2017).

Unfortunately, non-native accents often carry a social stigma (Gluszek and Dovidio, 2010). Because accents are one of the most immediate, powerful and fixed cues to one's cultural identity (Giles, 1977), they can reinforce and maintain stereotypes and prejudices between groups of people (Kinzler et al., 2007). In addition, they can be used as salient markers of socio-economic class and educational levels, which can lead native speakers to have a sense of superiority or inferiority compared to non-native accented speakers (Lippi-Green, 2012). Lippi-Green points out that this social hierarchy is so powerful that even the medical community treats the elimination of accents as an explicit goal in certain practices of speech therapy. Because native speakers and non-native speakers interact with one another more than ever (Cheng, 1999; Pickering, 2006), this leads to important questions about how this stigma plays out in social interactions and judgments within cross-cultural contexts.

Research investigating the perceptions and impressions of non-native accented speech has repeatedly shown that it is perceived less favorably than native accented speech on measures of believability (Lev-Ari and Keysar, 2010) and social preference (Kinzler et al., 2009, 2011; DeJesus et al., 2017). For example, Lev-Ari and Keysar (2010) found that people judged statements delivered by non-native accented speakers as less believable than when delivered by native accented speakers. In another study, social preference was measured by asking 5-year-old children to evaluate the likelihood of becoming friends with other children (Kinzler et al., 2009). The study found that, while American children chose the pictures of children with the same race when they were presented silently, they chose the pictures of children with the different race over those with the same race when the latter was speaking in French-accented English. Moreover, in a study that controlled for comprehensibility of non-native accents by using nonsense speech, researchers found that preschool-aged children sought and endorsed information from native accented speakers over non-native accented speakers (Kinzler et al., 2011). Because they used nonsense speech, this study revealed that comprehensibility was not a factor in the children's choices; rather, the preference was driven solely by the sound of the speech itself. Together, these studies show that speaking with a non-native accent comes at a significant social cost.

Hand Gestures and Native Language (L1)

Research has largely focused on how native and non-native accents interact with other cues to identity, like the race of the speaker (e.g., Rubin, 1992; Kinzler et al., 2011; DeJesus et al., 2017; Hansen et al., 2017). However, there is room for more research in the fluid aspects of communication that accompany accented speech, such as bodies, hands, and facial expressions that are a ubiquitous context when people speak (Kendon, 2004). For example, co-speech hand gesture—the natural movements of the hands and arms to co-construct meaning—is an essential component of everyday communication, so much so that some have theorized it should be treated as an integral part of language itself (McNeill, 1985, 1992, 2006). This fusion between speech and gesture justifies the importance of researching the two *together* when investigating all aspects of speech communication.

The integrated relationship between speech and gesture in language production has led many researchers to study how these two parts of the system work together during language comprehension (for reviews, see Hostetter, 2011; Kelly, 2017). Specifically testing McNeill's theory, Kelly et al. (2010) advanced the integrated systems hypothesis to show that that the semantic relationship between speech and gesture affect the accuracy and speed of language comprehension. Moreover, this semantic contribution appears to be bi-directional—gesture not only clarifies the meaning of speech, but speech itself clarifies the meaning of gesture. This tight relationship between speech and gesture has been further bolstered by research showing that speech and gesture are semantically integrated in traditional language networks in the brain (Willems et al., 2007; Wu and Coulson, 2007; Dick et al., 2009; Green et al., 2009; Holle et al., 2010).

Beyond semantics, co-speech gesture also serves a lower-level perceptual function as well. Indeed, researchers have shown that hand movements play a role in motor and acoustic processes, such as vocal production (Pouw et al., 2020) and prosodic accentuation (Krahmer and Swerts, 2007). For example, Krahmer and Swerts (2007) found that producing beat gestures with speech not only enhances acoustic properties of speech production, but they also help listeners perceive words to be more acoustically prominent in sentences, even when only the audio is presented. Moreover, when viewing beats, these gestures serve to enhance how viewers perceive prosodic stress in speech. On the neural level, this perceptual focusing function of gesture is evident in neuroimaging research showing that there tight coupling of gesture and speech during early stages of speech processing (Dick et al., 2009; Hubbard et al., 2009; Biau and Soto-Faraco, 2013; Wang and Chu, 2013; Skipper, 2014). In one early study, Hubbard et al. (2009) investigated the relationship between gesture and speech in the auditory cortex and found that compared to "speech with a still body" and "speech with nonsense hand gesture," speech accompanied by a congruent gesture elicited greater activation of auditory areas in the brain, such as the left hemisphere primary auditory cortex and the planum temporale (see also Dick et al., 2009).

This tight connection between viewing the hands and perceiving speech make gestures a useful tool in "speechreading," the ability to use visual cues of speakers to clarify what they are saying. In a pioneering (and under-cited) study, Popelka and Berger (1971) investigated how phrases presented in varying gesture conditions—ranging from no gesture to semantically congruent and incongruent iconic and deictic gestures—affected accurate perception of spoken sentences. They found that sentences presented with congruent gestures produced higher accuracy for hearing a spoken sentence than did sentences presented with no gestures, and both produced better accuracy than sentences accompanied by incongruent gestures. More recently, Drijvers and Özyürek (2017) discovered that when the auditory information is degraded, listeners particularly benefit from iconic gestures during speech comprehension (for similar evidence with people who are hard of hearing, Obermeier et al., 2011, or with "cued speech" representing the individual sounds of words with hands, LaSasso et al., 2003). However, when auditory information is too degraded, the "additive effect" from hand gestures is lost. So, it appears that not only do co-speech gestures help with understanding the meaning of an utterance, they also facilitate lower levels perceptual identification of the speech stream itself.

Hand Gestures and Second Language (L2)

Hand gestures are just as much part of using an L2 as they are using an L1 (Neu, 1990; Gullberg, 2006; McCafferty and Stam, 2009). Indeed, Gullberg argues that, given the integrated relationship between speech and co-speech gestures, the latter should be viewed as a fundamental part of the L2 elements that learners must master when acquiring an L2. Just as there are proper ways to phonetically articulate L2 syllables and syntactically organize L2 sentences, there seem to be fitting ways to move the hands when speaking a different language (Kita, 2009; Özyürek, 2017). This appropriate use of gesture applies to more than just the nuts and bolts of L2 phonetics, vocabulary and grammar—it also has pragmatic and cultural functions. In Gullberg's own words, "[t]he command of the gestural repertoire of a language is important to the individual learners' communicative efficiency and 'cultural fluency' (Poyatos, 1983)—perhaps less in terms of misunderstandings (Schneller, 1988) than in terms of the general integration in the target culture" (Gullberg, 2006, p. 116).

Many of the experiments on this topic have focused on how L2 learners attend to information conveyed through the hands when perceiving novel speech sounds (Hannah et al., 2017; Kelly, 2017; Kushch et al., 2018; Baills et al., 2019; Hoetjes et al., 2019) and comprehending new vocabulary (Allen, 1995; Sueyoshi and Hardison, 2005; Sime, 2006; Kelly et al., 2009; Morett, 2014; Morett and Chang, 2015; Baills et al., 2019; Huang et al., 2019). For example, Kelly et al. (2009) investigated how semantic congruence of gesture and speech affected the learning of L2 Japanese vocabulary in native English speakers. Results from a free recall and recognition test showed that compared to speech alone, congruent gestures enhanced memory and incongruent gesture disrupted it (and see Hannah et al., 2017, for a similar effect in L2 phonetic processing). Based on research in this vein, Macedonia (2014) makes a strong case for why hand gestures should be a bigger part of the L2 classroom and language education more generally.

But what about the other side of the coin? How do gestures produced by L2 speakers *themselves* affect native speaker's perceptions and impressions of those L2 speakers? There are a few notable studies that have addressed this question (Neu, 1990; Gullberg, 1998; Jungheim, 2001; Gregersen, 2005; McCafferty and Stam, 2009). For example, Gullberg (1998) observed that the more L2 learners produced co-speech gestures—particularly, iconic gestures—the more native speakers judged them to be generally proficient in the L2. This fits well with L1 research showing that co-speech gestures positively influence social evaluations of native speakers (Maricchiolo et al., 2009). And there is even some recent evidence that training L2 speakers to use co-speech gesture not only enhances impressions of those speakers, but also how those speakers actually produce L2 speech (Gluhareva and Prieto, 2017; Zheng et al., 2018; Hoetjes et al., 2019). For example, Gluhareva and Prieto showed that when native Catalan speakers were given training on how to pronounce English words with beat gestures, their L2 speech was judged by native English speakers to have improved significantly compared to when there was no training with beat gestures. Note that native speakers' judgments were on L2 speech alone, where they did not see learners' gestures. Thus, it remains to be seen if *viewing* L2 gestures affects how native speakers process lower level auditory aspects of L2 speech, such as, correctly hearing what was said or explicitly evaluating the non-native accent itself. In other words, it is possible that seeing L2 gestures not only helps to boost native speakers' social impressions of an L2 learner, it may also help them make better sense out of what they are hearing.

The Present Study

The present study explores this issue by focusing on a type of gesture that plays a powerful role in cross-cultural communication: emblematic gestures. Emblems are conventionalized movements of the hands, head and body that are understood by most members of one culture (or subculture), but not necessarily another (Efron, 1941; Ekman, 1972; Kendon, 1997; Kita, 2009; Matsumoto and Hwang, 2013). For example, in Japan, the emblem for, 'It's spicy,' is to hold the bridge of the nose with the thumb and index finger. Without culinary knowledge that wasabi causes a (strangely satisfying) burning sensation in the sinuses, this gesture would be quite baffling.

Emblems are interesting in an L2 context for a number of reasons. For one, they can be used simultaneously with L2 speech to create multimodal signals, and this allows L2 speakers to display additional knowledge about the L2 culture (Neu, 1990; Jungheim, 2001; Gullberg, 2006; Matsumoto and Hwang, 2013). Second, even though emblems are similar to words in that both have highly conventionalized forms, most emblems are less arbitrary than spoken words and exhibit an element of iconicity that more directly maps onto their cultural meaning (as with the "spicy" example) (McNeill, 1992; Poggi, 2008).[1] This gives L2 speakers an additional opportunity to convey meaning (similar to co-speech iconic gestures), which is particularly useful if their pronunciation is below the native level. And

[1] Of course, not all emblems have clear iconic meanings (e.g., the thumbs up gesture and OK sign mean different things across different cultures). Moreover, in other cases, the original iconic meanings become more obscure over time (e.g., it is believed that crossing the fingers for good luck is a vestigial iconic reference to the Christian cross) (Matsumoto and Hwang, 2013).

third, compared to the phonological challenges of L2 speech, emblems are relatively simple and easy to learn, making these visual conventions very handy in cross-cultural communication (Matsumoto and Hwang, 2013).

Emblems have not received much attention in the study of L1 speech comprehension, likely because they often occur independently of speech (Goldin-Meadow, 1999). However, in an L2 context, speakers can intentionally use culture-specific emblems along with speech to supplement the meaning of their utterances, in addition to demonstrating their sensitivity and knowledge of the L2 culture. Because viewing co-speech emblems helps L2 speakers comprehend L2 utterances (Allen, 1995), it is likely that they also help L1 speakers understand the non-native speech of L2 learners.

Building on this previous work, we ask the following question: From the perspective of native speakers, how does the cultural familiarity of L2 emblems affect phonetic perception of non-native accented speech specifically, in addition to the more general social evaluation of non-native speakers? This work extends the literature in three ways. First, previous studies on the perceptual processing and social stigma of accent (e.g., Gluszek and Dovidio, 2010; Lev-Ari and Keysar, 2010; Lippi-Green, 2012) have largely excluded its natural multimodal communicative context. If appropriately using hand gestures is an integral part of learning a complete L2 repertoire, as Gullberg (2006) argues, it makes sense to expand the focus and study non-native accents in their fully embodied form. Second, because many emblematic gestures are based on distinct and learned conventions—which often vary by culture—it is possible to explore the consequences of L2 speakers producing culturally right or wrong emblems. Just as a gesture's iconic meaning matters for L2 vocabulary learning (Kelly et al., 2009), it is possible a gesture's *cultural* meaning matters for perceptions and evaluations of L2 speech. Third, although previous research has shown that producing co-speech gestures in an L2 can make a general positive impression on native speakers—for example, Gullberg (1998) showing that gestures make L2 speakers appear more proficient—no study to our knowledge has more specifically broken down how L2 hand gestures influence the processing of non-native accents *per se* separately from the influence of gesture on evaluations of learners themselves.

In two experiments in two different languages, Mandarin and Japanese, we investigate how different gesture-speech relationships affect the evaluation of foreign language accent and learner from the perspective of native speakers. Specifically, we created gesture-speech pairs in which emblems that accompanied L2 speech were either culturally *familiar* or *unfamiliar* to native Mandarin or Japanese speakers. For both experiments, L2 learners were filmed performing a short utterance in three different conditions: culturally familiar gesture (common in China or Japan), culturally unfamiliar gesture (uncommon in China or Japan), and speech alone. In a within-subjects design, native Mandarin speakers watched videos (across all conditions) of L2 Mandarin learners, and native Japanese participants watched videos of L2 Japanese learners. Following each video, participants were asked a set of questions targeting speech perception and social impressions of the L2 learners.

We made two predictions about how L2 learners' gesture would affect L1 listeners' perception of speech and social impressions.

(1) We predicted that, relative to speech alone, culturally familiar gestures would improve accuracy and foreign accent ratings of L2 speech, and would positively affect social impressions of the accented speaker.

(2) In contrast, culturally unfamiliar gestures, relative to speech alone, would decrease accuracy and foreign accent ratings of L2 speech, and would negatively affect social impressions of the accented speaker.

These predictions were based on the following two lines of research as summarized in the introduction: one line of research showing that the relationship of gestures to speech matters for phonetic and semantic comprehension in L1 (Popelka and Berger, 1971; McNeill et al., 1994; Kelly et al., 2010) and L2 (Kelly et al., 2009; Hannah et al., 2017), and another line of research showing that the presence of meaningful gestures helps manage social impressions of L1 (Maricchiolo et al., 2009) and L2 speakers (Gullberg, 1998; Gregersen, 2005; McCafferty and Stam, 2009).

EXPERIMENT 1: MANDARIN

Methods

Participants

Thirty-six undergraduates (13 males and 23 females) from a small liberal arts university on the East Coast participated in Experiment 1. All participants were international students from different regions of mainland China. They were all judged by one of the authors from Beijing to be native Mandarin Chinese speakers. All of them learned English in school in China, and none grew up speaking it at home. None of their first exposure to English in the U.S. is earlier than age 15, but they scored 100 or higher in Test of English as a Foreign Language (TOEFL) at the time of admission to college. Participants received either academic credit in psychology or $5 in cash for their participation.

Materials

L2 learner stimuli

For the L2 learner video stimuli, we recruited twenty-one (14 males and 7 females) "learners" of Mandarin, who were students attending a small liberal arts university on the U.S. East Coast. None of the learners were native Mandarin speakers, and included a range of speaking Mandarin for the first time to those in intermediate and advanced level courses. The range of L2 Mandarin competency was intended to reflect varying levels of the Mandarin accent. Additionally, stimulus learners represented a wide range of racial, ethnic and gender diversity.

Video clips

The stimuli in the experiment consisted of twenty-one 2–4 s videos of Mandarin phrases that are common in everyday speech (see **Appendix 1 Supplementary Materials**). Each phrase was produced by a different learner in three conditions: (a) Speech + Culturally Familiar Gesture, (b) Speech + Culturally

Unfamiliar Gesture, and (c) Speech Alone. The "culturally familiar" gesture was defined as emblems that were familiar and commonly understood within Northern Mainland China. For example, consider the Mandarin utterance, "对不起 duì buqǐ," which means "Sorry." The culturally familiar emblem that goes with that speech is both palms meeting below the chin of the speaker, as in the left panel of **Figure 1**. A list of culturally familiar gestures was created with the help of native Mandarin speakers and Gestpedia[2], a website that documents gestures from various locations and cultures. To generate "culturally unfamiliar" gestures, we also consulted Gestpedia to find emblems for our various phrases that were associated with various different cultures. Some of these were taken from American culture, but some other cultures include Japanese, Nigerian, Vietnamese, and Egyptian. For example, a culturally unfamiliar gesture to native Mandarin speakers was a palm touching the chest, as in the middle panel of **Figure 1**. After an extensive list was compiled, the gestures were screened by three separate native Mandarin speakers to assure cultural familiarity.

During the recording phase, one of the authors, whose L1 is Mandarin Chinese, was present to ensure that learners' pronunciation of their assigned phrases were correct enough as to not to accidentally say a different word or phrase. Each learner said only one phrase but repeated it in the three conditions—familiar gesture, unfamiliar gesture, and speech alone—and all were videotaped. The stimulus clips were edited in Final Cut Pro and background noise in the audio clips was reduced with Audacity. In addition, the video clips were edited to have the same speech across all three conditions. To do this, the audio from the speech alone condition was dubbed onto all the other two versions of a given video to equate the speech across all conditions. This was important because it is known that producing hand gestures affects vocal production (Krahmer and Swerts, 2007; Pouw et al., 2020). Equally important was the naturalness of the audio and visual coupling. For this, we tested three people who were naïve to this experiment, and found that the stimuli all looked natural, and none of them noticed the dubbing. In summary, we created a total of 63 video clips (21 speakers × 3 conditions).

To prepare for the actual presentations of these video clips, three versions were created (see **Appendix 2 Supplementary Materials**), with the intention that each native speaker participant would take only one version of the experiment. Version A, B, or C each included all of the 21 learners, which meant that each version included all of the 21 utterances. But within each version, a learner appeared only in one of the three conditions. The condition in which the learner appeared was counterbalanced across the three versions. This was necessary to ensure that utterance type and gesture condition were not confounded, which is particularly important because there was a large range of accents across learners. In this way, we can control for diversity of accents by having each learner serve as his or her own control.

Evaluation of Learners' Videos

A set of eight questions was used in the questionnaire. They were grouped into two general categories of evaluation: (1) questions that measured perception of speech itself (*speech* evaluation) and (2) questions about social impressions of the Mandarin learners (*learner* evaluation).

Speech evaluation

To measure various forms of speech perception, the following questions were presented in Mandarin Chinese, which was the participants' L1: (1) *Words Misheard:* "What did this person say?" (fill in the blank); (2) *Accent:* "How would you rate their accent?" (1 = completely foreign to 10 = completely native Mandarin); and (3) *Tone Accuracy:* "How would you rate their tonal pronunciation?" (1 = completely incorrect to 10 = completely correct). The third question was specific to Mandarin as a tonal language, as it is possible to mispronounce a word in Mandarin by confusing one of the four lexical tones. In addition, we gave participants (4) a *Surprise Memory Test* at the end of the experiment, asking them to write down any of the learner's utterances that they could recall from the video. This was included because past research has shown that iconic gestures help disambiguate audio-degraded speech (Obermeier et al., 2011; Drijvers and Özyürek, 2017), and it is possible that this disambiguation would manifest in recall for accented speech too.

Learner evaluation

To probe for different aspects of social impression about L2 learners, the following questions were presented: (5) *Confidence:*

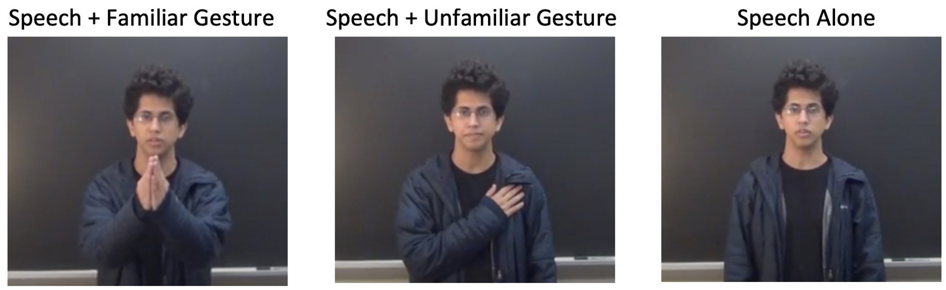


FIGURE 1 | Stimuli example from Experiment 1: Sorry (对不起 duì bu qǐ).

[2] https://www.gestpedia.com/

"How confident was this person?" (1 = not at all confident to 10 = extremely confident); (6) *Nervousness:* "How nervous was this person?" (1 = not at all nervous to 10 = extremely nervous); (7) *Communicative Effectiveness:* "How effective would this person be at communicating with native Mandarin Chinese speakers?" (1 = not at all effective to 10 = extremely effective); and (8) *Length of Study Time:* "How long do you think this person has been learning and practicing Mandarin Chinese?" [sliding scale labeled "amount of time (years)" from 0 to 20; it was converted to months later to be consistent with Experiment 2].

Procedure

The participants arrived at the Center for Language and Brain lab and were given a consent form. After they read the form, we clarified any questions before they signed it. The following script was read to participants of Experiment 1: "The purpose of this research is to evaluate the effectiveness of people speaking in Mandarin Chinese. You will view a series of brief videos of students practicing Mandarin Chinese, and after each one, take a survey to evaluate their learning efforts." The intention of this introduction was to prime the participants to treat the L2 learners in the stimulus video as students, in addition to getting participants in the mindset of providing constructive feedback to L2 speakers.

After the basic introduction of the task, the researchers encouraged participants not to spend more than 1 min responding to all eight of the video's questions. This time limit was introduced to emulate natural face-to-face communication in everyday life, during which listeners only have a very short time to process and integrate various sources of information about phonology, semantics, syntax, and pragmatics (Hanulíková et al., 2012). Participants were then brought into individual testing rooms, each containing of a computer, monitor and Pinyin keyboard on a desk.

The study was presented on Qualtrics. Participants were shown one video at a time, with each repeated twice. After that, the video would disappear from the screen, the set of seven survey questions appeared. The order of the questions, as described in the previous section, was set to a random order, and all participants answered them in this sequence: questions (5), (1), (2), (3), (8), (7), and (6). The experiment was self-paced, so the inter-stimulus interval length varied between participants. Each video and set of questions required about 45 s to 1 min. After participants finished responding to all the video stimuli, they were given the surprise memory test (question 4). The entire experiment lasted approximately 20–25 min.

After participants completed all of the tasks, the researcher debriefed them on the purpose of the study and compensated them with either course credit or $5 in cash.

Coding and Design

Aside from the rating scales, there were two measures that required coding: *Words Misheard,* with the question asking, "What did this person say," and the *Surprise Memory Test* at the end.

The *Words Misheard* question was coded by comparing the participant's typed answer to the actual speech in the video. A correct answer received a score of 0 (no errors), and an incorrect answer in any part of the utterance received a score of 1. The *Surprise Memory Test* involved free recall, and a score of "1" was given to phrases identical to the words presented in the study (complete memory) and a score of "0.5" was given to partially correct scores (partial memory), such as having the same root word but incorrect ending. A "0" was given for items that were entirely omitted or could not be traced back to any utterance (incomplete memory). In this way, low values for the "misheard" dependent variable (DV) mean better perception, whereas low values for the "memory" DV mean worse recall.

The experiment had a one-factor analysis of variance, with 3 conditions: culturally familiar gesture, culturally unfamiliar gesture, and speech alone.[3] Because we make non-orthogonal comparisons among our three levels of condition, we used Dunn-Šidák multiple contrasts to correct for Type I errors.

The DVs were separated into two categories. First, the L2 "Speech" evaluation includes measurements concerning (1) Words Misheard, (2) Accent, (3) Tone Accuracy, and (4) Memory Test. Second, the L2 "Learner" evaluation includes measurements concerning (5) Confidence, (6) Nervousness, (7) Communicative Effectiveness, and (8) Judgments of Length of Time Studying Mandarin.

Results

Speech Evaluation

Means and standard deviations of native Mandarin speaker responses are shown in **Table 1**. See the top half for the Mandarin data (see section "Experiment 1: Mandarin results).

For the proportion of misheard speech, there was a significant effect of gesture, $F(2,70) = 5.065$, $p = 0.014$, $\eta_p^2 = 0.16$. Familiar gestures produced lower error rates than both speech alone, tDS(3,35) = 2.757, $p = 0.014$, and culturally unfamiliar gestures tDS(3,35) = 2.743, $p = 0.030$. No significant difference was found between speech alone and unfamiliar gestures, tDS(3,35) = 1.03, n.s. The left panel of **Figure 2** shows the number of Mandarin words misheard in each of the three conditions (out of a total number of 756 answers = 21 utterances × 36 native listeners). The figure clearly demonstrates that the familiar gesture condition yielded the smallest number of misheard words, contrasting with the unfamiliar gesture and speech alone conditions.

On the evaluation of accent, there was a significant effect of gesture, $F(2,70) = 5.830$, $p = 0.005$, $\eta_p^2 = 0.143$. Familiar gestures produced significantly more native-like ratings compared to speech alone, tDS(3,35) = 3.061, $p = 0.006$, and also compared to unfamiliar gestures, tDS(3,35) = 2.776, $p = 0.014$. However, there was no significant difference between unfamiliar gestures and speech alone, tDS(3,35) = 0.281, n.s. For tonal accuracy, there was a significant effect of gesture, $F(2,70) = 4.206$, $p = 0.019$, $\eta_p^2 = 0.107$. Familiar gestures influenced participants to attribute more correct tonal pronunciation than speech alone, tDS(3,35) = 2.791, $p = 0.012$. However, there were no significant

[3] All of the analyses presented in both experiments used subjects as the error term. Ideally, we would have liked to also run parallel ANOVAs with item as the error term, but because we had far fewer items than subjects, our design was too underpowered to draw valid conclusions for item analyses. We address this limitation in the section "General Discussion."

TABLE 1 | Means and standard deviations of native speaker responses in the L2 "Speech" evaluation.

	Words Misheard		**Accent**		**Tone Accuracy**		**Memory Test**	
	1 = misheard 0 = correct		**10 = most native-like**		**10 = completely correct**		**100% = all words recalled**	
	Mean	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
Experiment 1: Mandarin								
FAMILIAR gesture	0.04	0.09	5.62	1.19	6.60	1.12	42.4%	1.08
UNFAMILIAR gesture	0.09	0.10	5.21	1.24	6.25	1.31	44.9%	1.20
SPEECH alone	0.11	0.10	5.17	1.28	6.14	1.12	33.3%	1.10
Experiment 2: Japanese								
FAMILIAR gesture	0.01	0.04	5.03	1.19	N/A	N/A	17.9%	0.14
UNFAMILIAR gesture	0.02	0.05	4.83	1.31	N/A	N/A	17.8%	0.13
SPEECH alone	0.06	0.07	4.67	1.30	N/A	N/A	17.3%	0.12

differences between familiar gestures and unfamiliar gestures, tDS(3,35) = 2.085, n.s., or between unfamiliar and speech alone, tDS(3,35) = 0.670, n.s.

The surprise memory test also yielded a significant effect of gesture, $F(2,70) = 5.045$, $p = 0.011$, $\eta_p^2 = 0.126$, such that speech alone yielded worse recall than both culturally familiar, tDS(3,35) = 2.500, $p = 0.026$, and unfamiliar gestures, tDS(3,35) = 3.332, $p = 0.006$. However, there was no significant difference between familiar and unfamiliar gestures, tDS(3,35) = 0.552, n.s.

Learner Evaluation

Means and standard deviations of native speaker responses in the L2 "Learner" evaluations were given in the upper half of **Table 2** (see section "Experiment 1: Mandarin results").

For confidence, there was a significant effect of gesture, $F(2,70) = 4.859$, $p = 0.011$, $\eta_p^2 = 0.122$, with speech alone lowering confidence ratings compared to both familiar, tDS(3,35) = 2.214, $p = 0.049$, and unfamiliar gestures, tDS(3,35) = 3.049, $p = 0.012$. There was no significant difference between the familiar and unfamiliar gestures, tDS(3,35) = 0.646, n.s. For nervousness, there was a significant effect of gesture, $F(2,70) = 3.311$, $p = 0.045$, $\eta_p^2 = 0.086$. The mean rating appeared higher, i.e., more nervous, in speech alone than in the other conditions, as shown in **Table 2**. However, none of the individual comparisons yielded a significant difference [familiar gestures vs. speech alone: tDS(3,35) = 2.159, n.s.; familiar vs. unfamiliar gestures: tDS(3,35) = 0.028, n.s.; and unfamiliar vs. speech alone: tDS(3,35) = 2.059, n.s.]. (Note that finding null results with our planned contrasts, despite finding a significant omnibus effect in the ANOVA, is the result of using Dunn-Šidák multiple contrasts, which adjusted the criteria more strictly than without an adjustment).

For communicative effectiveness, there was a significant effect of gesture, $F(2,70) = 6.644$, $p = 0.003$, $\eta_p^2 = 0.160$. Both familiar and unfamiliar gestures were judged to be more effective than speech alone [tDS(3,35) = 3.240, $p = 0.005$; tDS(3,35) = 2.619, $p = 0.039$, respectively]. Between familiar and unfamiliar gesture, however, no significant difference was found, tDS(3,35) = 1.388, n.s.

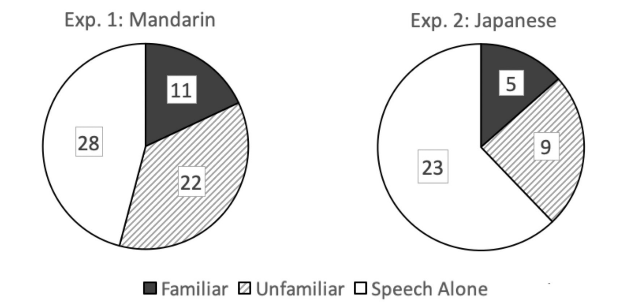


FIGURE 2 | Number of words misheard in each of the familiar-gesture, unfamiliar-gesture, and speech-alone conditions.

For estimates of time studying the Mandarin language, there was no significant effect of gesture, $F(2,70) = 1.457$, n.s.

Experiment 1 Summary

Speech evaluation

The most consistent finding in the speech evaluation measures was that familiar gestures indicated an advantage over speech alone in all dimensions: with fewer words misheard, higher "native-like" accent ratings, higher tone accuracy, and more recalled utterances in the surprised memory test (see **Table 3** for a summary of Experiment 1). However, effects of unfamiliar gestures were somewhere between the other two conditions—in two evaluations (tone accuracy and memory test), unfamiliar gestures did not differ from familiar gestures, but in the other two evaluations (words misheard and accent ratings) unfamiliar gestures showed significantly less advantage than familiar gestures. Compared with speech alone, unfamiliar gestures had only one advantage, producing more recalled items in the surprised memory test than speech alone, but they did not differ in the other evaluations. Our original prediction was that unfamiliar gestures would have a more negative effect than speech alone, but none of the cases showed this.

Learner evaluation

Two major patterns were found for evaluation of L2 learners. First, we found that familiar and unfamiliar gestures both led

TABLE 2 | Means and standard deviations of native speaker responses in the L2 "Learner" evaluation.

	Confidence		Nervousness		Comm. Effectiveness		Months Studying*	
	10 = extremely confident		10 = extremely nervous		10 = extremely effective		0–50 months	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Experiment 1: Mandarin								
FAMILIAR gesture	6.55	1.09	3.54	1.16	6.61	1.15	30.10	1.22
UNFAMILIAR gesture	6.65	1.19	3.55	1.09	6.42	1.15	29.67	1.23
SPEECH alone	6.17	1.29	3.91	1.16	6.13	1.17	27.67	1.38
Experiment 2: Japanese								
FAMILIAR gesture	7.15	1.05	2.96	1.42	6.84	1.03	15.72	6.66
UNFAMILIAR gesture	6.87	1.19	3.22	1.43	6.60	1.08	14.97	7.12
SPEECH alone	6.47	1.30	3.80	1.46	6.47	1.22	14.38	6.12

**The original data in 'years' for Mandarin experiment were converted to 'months' to match Japanese data.*

to higher ratings of confidence and communicative effectiveness, compared to speech alone. In contrast, for the evaluation of nervousness and the estimate of time studying Mandarin, there were no differences among the three conditions.

EXPERIMENT 2: JAPANESE

Experiment 2 attempted to build on Experiment 1 by generalizing to a different language and culture: Japanese. Given that the vast majority of research in psychology has focused on Western societies and English speakers, it is important to increase diversity in the field by expanding to different cultures and languages (Henrich et al., 2010). It goes without saying that there are vast differences among Asian languages and cultures as well. This diversity is especially relevant for the topic of emblematic gestures, which by definition depend on the specific conventions of a particular culture.

TABLE 3 | Summary of significant differences between conditions: FAMILIAR Gesture, UNFAMILIAR Gesture, and SPEECH Alone.

		FAMILIAR vs. SPEECH	FAMILIAR vs. UNFAMILIAR	UNFAMILIAR vs. SPEECH
(1) Speech evaluation				
Exp 1: Mandarin	Words misheard	*	*	n.s.
	Accent	**	*	n.s.
	Tone	*	n.s.	n.s.
	Memory test	*	n.s.	**
Exp 2: Japanese	Words misheard	***	n.s.	*
	Accent	**	n.s.	n.s.
	Memory test	n.s.	n.s.	n.s.
(2) Learner evaluation				
Exp 1: Mandarin	Confidence	*	n.s.	*
	Nervousness	n.s.	n.s.	n.s.
	Comm. Effectiveness	**	n.s.	*
	Months studying	n.s.	n.s.	n.s.
Exp 2: Japanese	Confidence	***	n.s.	***
	Nervousness	***	*	***
	Comm. Effectiveness	**	*	n.s.
	Months studying	*	n.s.	n.s.

**p < 0.05, **p < 0.01, ***p < 0.001. In all cases, the direction of differences was more positive evaluation (e.g., less words misheard, more words memorized, more confidence, and less nervousness) for FAMILIAR than SPEECH, FAMILIAR than UNFAMILIAR, and UNFAMILIAR than SPEECH.*

Combined with the authors' impressions and discussions with native Chinese and Japanese speakers, we reasoned that these two cultures might vary to different degrees in the use of gesture, making Japanese emblems a good candidate for the present study.

More importantly, we considered another point: The Mandarin speakers in the first experiment were enrolled in a university in the U.S. for 0.5–3.5 years at the time of testing, and they were proficient in English for undergraduate studies. This factor might have exposed them to a greater variety of linguistic and cultural elements outside their native language and culture, and it might have made them more open to difference than people who have never lived abroad. With this in mind, we sought to find college students in Japan who did not have as extensive experience abroad. This may make the interpretation of emblems relatively more uniform across these participants in Japan, which would be a nice contrast with the Chinese participants in Experiment 1.

Using the same basic paradigm as Experiment 1, we investigated the extent to which native Japanese speakers are sensitive to the cultural meaning of emblem gestures when: (1) perceiving non-native speech and (2) forming social impressions of non-native speakers.

Method

The method for Experiment 2 was largely borrowed from Experiment 1 with a few notable differences that will be addressed in this section.

Participants

Forty-eight native Japanese undergraduate college students (all females) from a small all-women's college in Tokyo participated in Experiment 2. All participants had limited exposure to the English language, mostly having learned it formally in school. None of them had experience of studying abroad for more than a year. Participants received 1,000 yen, the rough equivalent of ten U.S. dollars, for their participation.

Materials

L2 learner stimuli

There were thirty "learners of Japanese" who were students of a small liberal arts university on the East Coast in the U.S. Similar to Experiment 1, learners represented varying levels of accented Japanese speech. Learners ranged from no exposure to students who had been learning the Japanese language for 3 or 4 years (400-level). Learners on the video represented a wide spectrum of racial, ethnic and gender diversity.

Video clips

The process of creating the 30 video clips was the same as described in Experiment 1. Each stimulus was assigned a short Japanese phrase. For example, consider the Japanese phrase for saying "Go for a drink?," which was "Nomi ni ikanai?" In Japanese, a culturally familiar emblem for "Go for a drink?" is a two-finger gesture with the index finger and thumb positioned horizontally, tilting toward the mouth. To create the culturally unfamiliar condition, the same speech would be paired with a Russian gesture for "Go for a drink?": a tilt of the head with a light flick on the side of one's neck. As a baseline, the third condition was the learner speaking in the video without any gesture. Each learner was first instructed to perform in these three conditions and then videotaped. See **Figure 3**.

After all the videos were created, there were some concerns that a few of our gesture-speech pairs were not as good as others. To explore this possibility, we ran a "norming" test asking four native Japanese speakers to evaluate the cultural familiarity of our emblems in each of our familiar-unfamiliar pairs across all 30 items. Specifically, the four native Japanese speakers first read the Japanese phrase, and then viewed each of the gestures, familiar and unfamiliar, paired with that phrase. For each phrase, they were asked to keep it in mind and judge how naturally the gesture captured that meaning in the scale of 1 (not at all natural) to 10 (completely natural). Based on this norming study, it was discovered that two items had a pattern of rating in which the familiar and unfamiliar emblems were rated as very close to one another and two items in which the familiar gestures were rated as *less natural* than the unfamiliar gestures. Consequently, these items were eliminated from all analyses presented below. This removal did not change the significance of the results, except for the question about how long learners had been learning Japanese. The total of 26 stimuli used in the final analysis were shown in **Appendix 3 Supplementary Materials**.

Evaluation of Learners' Videos

The questionnaire was very similar to Experiment 1, with two sets of questions focusing on (1) speech and (2) social impressions of learners, except that it was given in Japanese, the participants' L1 in Experiment 2. There were four minor changes in the wording of a few of the questions. First, in Experiment 1, the scale representing accentedness read: 1 (completely foreign) to 10 (completely native). In Experiment 2, this was changed to: 1 (not at all native) to 10 (completely native) to maintain consistency within the vocabulary used. Second, another question in Experiment 1 asked the participants to estimate how long the learner "has been learning and practicing Mandarin Chinese" on a scale ranging from 1 to 20 years. This scale did not seem to be very effective, as the mean score for each condition displayed around 2 years of perceived learning. For Experiment 2, we altered the scale to 1 – 50+ months which was labeled on a sliding scale. Third, while in Experiment 1 the scale measuring nervousness read 1 (not at all nervous) to 10 (extremely nervous), Experiment 2's nervousness scale was inverted: 1 (extremely nervous) to 10 (not at all nervous). To be consistent with Experiment 1, we converted the scores from Experiment 2 to the scale in Experiment 1, in which higher numbers mean *more* nervous. And fourth, the question relating to tone in Experiment 1 was removed for Experiment 2, given the difference in the use of fundamental frequency in Japanese and Chinese phonology (Howie and Howie, 1976; Vance, 2008).

Procedure

The basic procedure was the same as Experiment 1, but the instructions were given in Japanese by one of the two experimenters who spoke advanced Japanese. The testing site was also different from Experiment 1 because Experiment 2 took place at a small all women's college in Japan. Time slots for the study were set up so that 2 participants would come for the study at the same time. The testing room was set up so that the tables lined the perimeter of the room. Participants sat in the two corners, each setup with a laptop and headphones, facing the same wall so that the researchers could see when they finished. The study took about 45 min to complete. The experimenters waited until both participants were done, and they were debriefed together in Japanese at the end.

Coding, Design, and Analyses

We used the same basic design as Experiment 1, which was a one-factor analysis of variance, with condition (3 levels) as a within-subjects factor. The open-ended questions (*words misheard* and *memory test*) were coded in the same way as Experiment 1.

Results

Means and standard deviations of native Japanese speaker responses are shown in **Tables 1**, **2**. See the bottom half of each table for the Japanese data.

Speech Evaluation

For the proportion of misheard speech, there was a significant effect of gesture, $F(2,94) = 8.076$, $p = 0.001$, $\eta_p^2 = 0.147$, with speech alone producing higher proportions of errors than both familiar gestures, tDS(3,47) = 3.766, $p < 0.001$, and unfamiliar gestures tDS(3,47) = 2.615, $p = 0.036$. There was no difference between unfamiliar and familiar gestures, t(3,47) = 1.077, n.s. The right panel of **Figure 2** shows the number of Japanese words misheard in each of the three conditions (out of a total number of 1,248 answers = 26 utterances × 48 native listener participants). Although the total number of misheard words was quite small, it is notable that roughly 60% of the errors occurred in the speech alone condition.

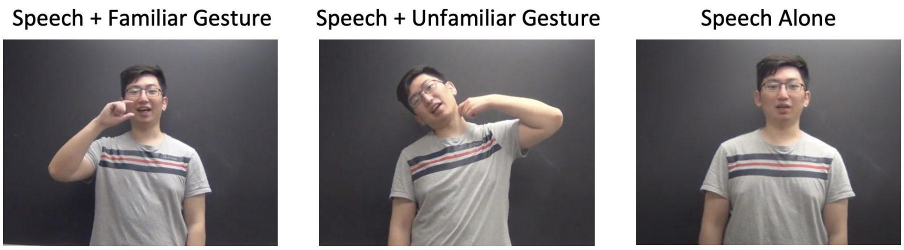


FIGURE 3 | Stimuli example from Experiment 2: Go for a drink? ("Nomi ni ikanai"?).

There was a significant effect of gesture on accent perception, $F(2,94) = 4.980$, $p = 0.010$, $\eta^2_p = 0.096$, with familiar gestures producing higher native-like ratings than speech alone, tDS(3,47) = 3.087, $p = 0.005$. However, unfamiliar gestures did not significantly differ from familiar gestures, tDS(3,47) = 1.978, n.s., and from speech alone, tDS(3,47) = 1.293, n.s.

For the surprise memory test, there was no significant effect of gesture, $F(2,94) = 0.033$, n.s.

Learner Evaluation

For confidence, there was a significant effect of gesture, $F(2,94) = 13.645$, $p < 0.001$, $\eta^2_p = 0.225$. Familiar gestures produced significantly higher confidence ratings than speech alone, tDS(3,47) = 4.690, $p < 0.001$. In addition, unfamiliar gestures produced higher scores than speech alone, tDS(3,47) = 3.672, $p < 0.001$. However, scores did not differ between familiar and unfamiliar gestures, tDS(3,47) = 2.045, n.s.

For nervousness, a significant effect of gesture was also found, $F(2,94) = 20.310$, $p < 0.001$, $\eta^2_p = 0.302$. Familiar gestures produced lower nervousness ratings than both speech alone, tDS(3,47) = 5.825, $p < 0.001$, and unfamiliar gestures, tDS(3,47) = 2.328, $p = 0.036$. In addition, unfamiliar gestures produced lower nervousness ratings than speech alone, $t(3,47) = 3.971$, $p < 0.001$.

For communicative effectiveness, there was also a significant effect of gesture, $F(2,94) = 5.725$, $p = 0.006$, $\eta^2_p = 0.109$, with familiar gestures judged as more effective than both speech alone, tDS(3,47) = 2.888, $p = 0.009$, and unfamiliar gestures, tDS(3,47) = 2.354, $p = 0.035$. However, there was no difference between unfamiliar gestures and speech alone, tDS(3,47) = 1.254, n.s. For estimates of time studying the Japanese language, a significant effect of gesture was also found, $F(2,94) = 3.146$, $p = 0.048$, $\eta^2_p = 0.063$. Learners with familiar gestures were judged as studying the language longer than those with speech alone, $t(3,47) = 2.456$, $p = 0.027$. However, no other comparisons yielded significant differences.

Experiment 2 Summary

Speech evaluation

Table 3 presents a summary of Experiment 2. Familiar gestures were associated with less mishearing and higher 'native-like' accent ratings than speech alone, showing their advantage. Interestingly, familiar gestures did not differ from unfamiliar gestures in both of these two evaluations. Unfamiliar gestures showed one advantage over speech alone, having less misheard words. Unlike Experiment 1, there were no differences across conditions in recall accuracy for the memory test.

Learner evaluation

Positive effects of familiar gestures were robust in the learner evaluation: Familiar gestures were associated with more confidence, less nervousness, more effectiveness in communication, and judgments of longer months of study than speech alone. In addition, familiar gestures were more advantageous than unfamiliar gestures in two of the four evaluations as well (less nervous and more effective in communication), but not in the other evaluations. Just as in Experiment 1, effects of unfamiliar gestures were somewhere between the other two conditions: Unfamiliar gestures produced higher confidence ratings and lower nervousness ratings than speech alone, but the two conditions did not differ in the other two evaluations.

GENERAL DISCUSSION

Culturally Familiar Gestures Help, Uniformly

The results from the two experiments, as summarized in **Table 3**, provide strong support for our first prediction. We predicted that, relative to speech alone, culturally familiar gestures would improve speech perception and memory, as well as social impressions of the L2 learner (Popelka and Berger, 1971; Gullberg, 1998; Gregersen, 2005; Kelly et al., 2009; Maricchiolo et al., 2009).

In Experiment 1, we found that familiar gestures produced more positive responses than speech alone in all of the *speech* evaluation dimensions: fewer perception errors, higher "native-like" accent ratings, higher tone accuracy, and greater words recalled in the surprise memory test. Similarly, Experiment 2 revealed that familiar gestures produced fewer perception errors and higher accent ratings compared to speech alone (but, unlike Experiment 1, such benefit was not observed in the memory test).

These advantages of familiar gestures over speech alone extend to include the social impression of L2 *learners*. Culturally familiar gestures raised ratings in two of the four evaluations—confidence and communicative effectiveness—in Experiment 1, and in addition, they positively affected all of the evaluations in Experiment 2, including the lower judgments of nervousness and higher estimates of how long learners had been studying

Japanese. The findings that familiar gestures positively influenced speech perception is consistent with literature showing that semantically related speech and gesture improve accuracy of L1 comprehension (Popelka and Berger, 1971; Graham and Argyle, 1975; Kelly et al., 2010; Dahl and Ludvigsen, 2014) and vocabulary retention in L2 learning (Allen, 1995; Sueyoshi and Hardison, 2005; Sime, 2006; Kelly et al., 2009; Morett, 2014), in addition to boosting speech perception when auditory information is moderately degraded (Obermeier et al., 2011; Drijvers and Özyürek, 2017). Adding to this work, the present study demonstrates that the *cultural* relationship between L2 speech and gesture matters, too. When gestures culturally match the L2—what we call, culturally familiar emblems—they play a positive role in shaping how L2 speech is perceived. Moreover, going beyond previous work by Allen (1995), our results show that not only is the mere presence of emblematic gestures useful, but their specific cultural content matters, too.

Focusing first on perception errors, what mechanism might explain why culturally familiar gestures best help native speakers to hear speech correctly? Considering Experiment 1, there were not many instances of misheard speech across the board (about 8%), but familiar gestures were particularly low with only a ∼4% error rate. In contrast, unfamiliar gestures more than doubled that rate (∼9%) and having no gestures produced even more errors (∼11%). In Experiment 1, familiar gestures also boosted judgments of Mandarin tonal pronunciation accuracy. One possibility is that because culturally familiar gestures are so easily recognizable for native speakers, it may have required minimal cognitive effort to process their meaning, leaving adequate perceptual resources to focus on the L2 speech (Adank et al., 2009).

With regard to accent ratings, the results from both experiments add to the literature on the phonological functions of co-speech gesture. While previous research has shown that the hand movements of speakers—in an L1 (Krahmer and Swerts, 2007; Pouw et al., 2020) and L2 (Gluhareva and Prieto, 2017; Zheng et al., 2018; Hoetjes et al., 2019)—affect perceptions of speech by L1 users, no study to our knowledge has shown that *viewing* culturally familiar gestures can modulate how non-native accents are perceived by native speakers. In both experiments, we show that the presence of culturally familiar gestures improves ratings of accentedness compared to no gestures. Previous research has shown that contextual factors, such as race of speaker (Jussim et al., 1987; Rubin, 1992; Hansen et al., 2017), can modulate perception of accent; here, we extend this phenomenon to include not just these fixed features of the context, as in the case of speaker identity, but more fluid factors, such as what people do with their hands. This fits well with research on the processing of speech in the context of other dynamic multimodal signals, such as the integration of facial expressions, body posture and emotional tone of voice (Pourtois et al., 2005; Van den Stock et al., 2007).

The benefits of producing culturally familiar gestures also extend to managing social impressions of others. Consistent with our first prediction, we found that, for both experiments, the presence of familiar gestures led to more positive impressions than speech alone. This work fits nicely with previous studies on the social benefits of co-speech gesture for L1 (Maricchiolo et al., 2009) and L2 speakers (Gullberg, 1998; Gregersen, 2005). For example, Gullberg (1998) found that native speakers made more positive evaluations of L2 learners who produced many iconic gestures. And with regard to nervousness, Gregersen (2005) showed that native speakers judged L2 learners to be more at ease when they used many emblematic and iconic gestures, and in contrast, more anxious when they produced mostly non-communicative self-adaptors (e.g., fidget with objects, touching face and hair, adjusting clothing) or no gestures at all (e.g., hands in lap or arms crossed). Moreover, all of these studies focused on Indo-European languages and learners from the US and Europe, and we extend beyond that by adding data from non-Indo-European languages, Mandarin and Japanese, and a different part of the world, Asia.

It is worth adding that in Experiment 2, culturally familiar gestures helped social impressions more than unfamiliar ones (**Table 3**). Specifically, familiar gestures lowered assessments of nervousness and improved judgments of communicative effectiveness for the Japanese viewers. This suggests that at least some of the time, simply waving the hands is not enough to make a good impression—the cultural meaning of gesture matters. It is interesting that this pattern held only for the native speakers of Japanese, but not Mandarin. One possible reason for this inconsistency is that the culturally familiar and unfamiliar gestures were better differentiated for Japanese native speakers in Experiment 2. This could be due to there being more consistency and uniformity of emblems in Japan and more variability of emblems in China. Another possibility is that because we tested Chinese international students—who were attending college abroad in the United States with other international students—it could be that they simply had been exposed to a wider a diversity of emblems. This cultural exposure may have made them more open-minded to "unfamiliar" gestures and ultimately diluted the difference between the two conditions. We will return to the differences between our two samples in a later section. It is also possible that norming familiar and unfamiliar emblems in Experiment 2, but not in Experiment 1, contributed to the different findings for social impressions between the two experiments.

Culturally Unfamiliar Gestures Help, Variably

Our second prediction was that, relative to speech alone, culturally *unfamiliar* gestures would decrease foreign accent ratings and encoding/memory accuracy of L2 speech (Popelka and Berger, 1971; McNeill et al., 1994; Kelly et al., 2009), in addition to lowering social impressions of the accented speaker. Our results in the two experiments indicated that this prediction was not supported at all. In no cases did unfamiliar gestures produce significantly more negative responses than speech alone in the evaluations of L2 speech and L2 learners. Instead, unfamiliar gestures produced more advantageous ratings than speech alone in some evaluation questions (the right-most column in **Table 3**). For example, compared to speech alone, unfamiliar gestures were associated with greater number of recalled words in the Mandarin experiment, and with fewer misheard words in the Japanese experiment, while other evaluations yielded no difference between the two conditions. For the social evaluation of L2 learners, unfamiliar gestures, compared to speech alone, were associated with more confidence and higher communicative effectiveness in the Mandarin study, and with more confidence and less nervousness in the Japanese study.

We also found a surprising result that, in the majority of evaluation questions, unfamiliar gestures did not differ from familiar gesture (see the middle column in **Table 3**). For example, none of the evaluation questions showed a difference between familiar and unfamiliar gestures in the social impression of the Mandarin learners, and also there were no differences in speech evaluations of the Japanese learners. This pattern was unexpected given that native speakers have difficulty processing non-native speech under adverse listening conditions (Adank et al., 2009; Bent and Atagi, 2017). From that perspective, we expected unfamiliar emblems to distract native speakers, depleting their perceptual resources, which would cause them to make more encoding errors than the optimal familiar gesture condition—but that was not the case. What might be going on?

A prominent framework for research on multimodal communication is Clark and Paivio's dual coding theory of information processing (Paivio, 1990; Clark and Paivio, 1991). By this traditional account, communication is enhanced when there is both a verbal and imagistic channel, and this is theorized to be the case even when the two channels do not convey the same semantic content. Although most gesture researchers treat the semantic relationship between speech and gesture as critical, there is some evidence that semantic congruence is not always essential. For example, even beat gestures, which often have little inherent semantic connection to speech, affect L1 speech processing (Krahmer and Swerts, 2007; Biau and Soto-Faraco, 2013; Wang and Chu, 2013) and memory (So et al., 2012). And in an L2 context, there is evidence that viewing and producing a range of hand movements—beat gestures (Kushch et al., 2018), metaphoric pitch gestures representing lexical tone (Morett and Chang, 2015; Baills et al., 2019) and even iconic gestures with idiosyncratic meanings (Macedonia and Klimesch, 2014; Huang et al., 2019)—can help with L2 vocabulary learning and retention. Connecting these findings to the present study, it is interesting that the presence of any gesture increased memory for speech in Experiment 1 (both gesture conditions produced a ∼30% improvement in recall over speech alone) and decreased the number of "misheard" utterances in Experiment 2 (both gesture conditions reduced errors by over 60% compared to speech alone). This suggest that at least on occasion, the mere act of moving the hands as a non-native speaker may help draw attention to the accompanying speech, much like a beat gesture functions, while also providing a visual anchor to help listeners remember what was said—no matter the meaning of the gesture.

For social impressions of the learners, the presence of any type of emblem also seemed to have some benefits. It is possible that our gestures, even when they culturally missed the mark, functioned to signal social effort, which may have led participants to evaluate learners who gestured in a more positive light (Gullberg, 1998; Maricchiolo et al., 2009). In the case of the Japanese experiment, perhaps not gesturing at all was a sign of anxiety when speaking the L2, whereas simply moving the hands to *intentionally* communicate anything—no matter whether it was culturally appropriate—signaled that L2 learners were more at ease (Gregersen, 2005).

Finally, it is worth considering the possibility that our participants did not always view our "culturally unfamiliar emblems" as emblems *per se*. Perhaps they occasionally viewed them as regular co-speech iconic gestures, albeit unusual and obscure ones. Because co-speech iconic gestures are not bound by conventional standards as much as emblems, and because many of the unfamiliar emblems in the two experiments had distinct iconic properties, native speakers may have given some of the unfamiliar gestures much more leeway when produced by L2 speakers. This highlights the important issue of variability, which we discuss next.

Variability

Traditionally, finding variability in results across samples and within conditions is seen as a red flag, an indication of weak external and internal validity. However, we see it differently in the present study. For one, collecting diverse samples of participants intentionally opens the door to more variability (Henrich et al., 2010). Beyond the diversity of studying non-native English speakers from Asia, we also had important differences between our two samples: In Experiment 1, we studied native Mandarin speakers who were fluent in English and attended college in the United States, whereas in Experiment 2, we studied native Japanese speakers who were mostly monolingual and had not spent extended periods outside of Japan. These differences are sure to cause some variation in the results.

Looking at the comparison between familiar and unfamiliar gestures in the learner evaluation (the middle column of **Table 3**), unfamiliar gestures were associated with disadvantage in none of the social impression questions for the Mandarin Chinese participants. This contrasts with the Japanese participants showing a disadvantage of unfamiliar gestures in two of the four social impression questions. One possibility is that participants in our Mandarin sample may have been exposed to a wider diversity of emblems, both in China with its diversity of cultural gestures (based on its higher linguistic diversity) and in America on a college campus with students and faculty from dozens of countries from around the world. This exposure might have contributed to Chinese participants more generously appreciating the speakers' effort than the Japanese participants' exposure mostly to their domestic gestures.

Another source of variability comes from the diverse functions of gestures themselves (Church et al., 2017; Novack and Goldin-Meadow, 2017). For example, Novack and Goldin-Meadow (2017) point out that gestures play multiple roles across contexts—communicating, problem solving, learning and remembering—and across social roles—for those who *view* gesture and those who *do* gesture. The present study taps into a wide range of these multiple functions: L2 learners produced emblems while communicating foreign utterances, and then native speakers viewed those gestures to perceive and recall speech, form judgments about non-native accents, and make social assessments about communicative effectiveness, confidence and nervousness. Given these varied functions, it makes sense that the cultural familiarity of gesture may at times be important in social categorization, but not in perceptual processing. And at other times, it is not surprising that it is the other way around.

Finally, one limitation of our study is that although it was well powered to run subject analyses, it was not adequately powered to run item analyses. Still, we did run unofficial item analyses on eleven of our fifteen dependent measures across both experiments. In our lower powered experiment (Experiment 1, which had 21 items), five of our significant effects were lost, but in our higher powered experiment (Experiment 2, which had 26 items), only one effect was lost. Interestingly, in each experiment, there was one new significant effect. Because these were underpowered analyses, it is hard to interpret them: On one hand, it could be that there indeed was more variability across items than subjects; but on the other hand, the variability could actually be comparable, but because there were far fewer items than subjects, the statistical differences among conditions was diluted in the item analyses. Following up on this, if increasing the number of items produces similarly robust effect sizes as the present study, it would strengthen our conclusion that co-speech emblems plays a beneficial role in cross-cultural contexts.

Future Studies

It is worth noting that there is another function of gesture that was intentionally missing from the present study, but likely would have also played a major role. Recall that producing gestures affects vocal production in an L1 (Krahmer and Swerts, 2007; Pouw et al., 2020) and L2 (Gluhareva and Prieto, 2017; Zheng et al., 2018; Hoetjes et al., 2019). In the present study, we dubbed identical speech onto each of our three video conditions in order to control for this vocal function of gesture. However, in the wild, this vocal effect of gesture runs free. This means that there may be layered roles of cultural emblems: not only would they function to visually influence the way spoken information is processed and evaluated by others (as we have shown), but they may also directly affect the quality of the actual speech signal itself. Going forward, it would be interesting to move beyond showing that culturally appropriate gestures positively influence how non-native speech is *received* and also explore whether a gesture's cultural appropriateness affects how non-native speech is actually *produced*. Does asking, "Nomi ni ikanai?," with the *right* drinking emblem help a learner vocally articulate that Japanese utterance any better? This is an interesting question to pursue in the future.

Even if producing appropriate emblems does not actually help learners pronounce L2 speech, it could make them *believe* it does. Consider that in a recent study by Zheng et al. (2018), novice L2 speakers of Mandarin self-reported that making the gestures corresponding to lexical tones was vastly more helpful in pronouncing the tones than not gesturing at all. And anecdotally, during the filming session of our study, many of the L2 learners informally commented that their pronunciation felt the best when they produced gestures. In this way, producing emblems—culturally right or wrong—may serve multiple and varied purposes in cross-cultural communication, and future work should attempt to disentangle these diverse functions.

Theoretical and Practical Implications

Starting with David McNeill's seminal 1985 paper, *So you think gestures are nonverbal*, there has been growing interest in understanding gesture and speech as an integrated semiotic system, as a window into the mind of a speaker. Indeed, many of the papers in this *Frontiers Research Topic* are focused on mental aspects of this integrated system of meaning. However, as Kendon (2017) recently pointed out, this focus on the cognitive components of gesture—while extremely valuable—has often eclipsed the many potent social function of the hands (see also Church et al., 2017). Kendon reminds us that gestures also have a distinct cultural component (Kendon, 1997), and together with speech, the two modalities combine to create a powerful pragmatic tool (Kendon, 2017). And if this is the case for one's L1, it may apply doubly for wielding a second language. Recall that Gullberg (2006) makes a strong case that mastering an L2 gesture repertoire is key to a learner's "cultural fluency." Indeed, as far back as Efron (1941), we have known that gestures signal social identity and that learning to adapt them to new contexts and environments is a sign of successful cultural assimilation.

This cultural component of hand gesture has been absent in research on the social stigma of non-native accents (Giles, 1977; Gluszek and Dovidio, 2010; Lev-Ari and Keysar, 2010; Kinzler et al., 2011; Lippi-Green, 2012; DeJesus et al., 2017). This is noteworthy because although relatively fixed aspects of one's identity (e.g., gender, race and class) are well known to affect how accents are received (Jussim et al., 1987; Rubin, 1992; Van Berkum et al., 2008; Hansen et al., 2017), there has been much less attention to how more fleeting aspects of context influence accent perception and evaluation. What speakers do with their bodies is a ubiquitous, but fluid and ever-changing, part of the way one speaks a native or non-native language. Focusing specifically on non-native accents, we have shown that this dynamic gestural context can affect many different aspects of how native speakers receive accented speech: correctly or incorrectly hearing and remembering what was said; positively or negatively shifting evaluations of pronunciation; increasing or decreasing impressions of confidence and nervousness; and raising or lowering judgments of communicative and cultural competence.

Bridging these two lines of work—research on accent and research on gesture—opens up new and important practical and theoretical questions. For example, how does what you do with your hands interact with more stable features of one's identity, like race, gender or class? Because non-native accents are so hard to change beyond the sensitive period (Johnson and Newport, 1989), might gesture be used as a compensatory tool to give speech a hand? Given that traditional L2 instruction typically focuses on teaching correct *spoken* language (Jungheim, 2001), could this instruction be improved by also teaching students how to correctly gesture more systematically and comprehensively? This is an exciting question since it may bear on a learnable element that gives everyone a chance to improve, contrasting with one's fixed social identity or hard to change non-native accent.

CONCLUSION

To our knowledge, no previous study has explored the combined perceptual and social benefits of co-speech emblems in L2 communication. The results from our two experiments suggest that, during cross-cultural communication, visual information conveyed through hand gesture influences low level phonetic perception, in addition to higher level social evaluation. We have shown that perception and evaluation improve when L2 speakers use emblems—both culturally familiar and unfamiliar—even if non-native accents themselves stay the same and even when it spans very short utterances of a few seconds. This suggests that in cross-cultural communication, more attention should be paid to what L2 learners do with their hands.

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 the Colgate University, Internal Review Board. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SK, YH, and KB-V: question formation. KB-V and ZL: stimulus creation and data collection. SK, YH, KB-V, and ZL: coding, transcription, and analysis. KB-V: writing first draft. SK and YH: writing later drafts. YH: figures and tables. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Picker Interdisciplinary Science Institute of Colgate University.

ACKNOWLEDGMENTS

We thank Chau Le for taking the lead on the Mandarin study, and thank student and alumni members of the Center for Language and Brain at Colgate for their helpful input.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.574418/full#supplementary-material

REFERENCES

Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 520–529. doi: 10.1037/a0013552

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *Modern Lang. J.* 79, 521–529. doi: 10.2307/330004

Atagi, E., and Bent, T. (2017). Nonnative accent discrimination with words and sentences. *Phonetica* 74, 173–191. doi: 10.1159/000452956

Baills, F., Suárez-González, N., González-Fuente, S., and Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of mandarin Chinese tones and words. *Stud. Sec. Lang. Acqu.* 41, 33–58. doi: 10.1017/S0272263118000074

Bent, T., and Atagi, E. (2017). Perception of nonnative-accented sentences by 5- to 8-year-olds and adults: the role of phonological processing skills. *Lang. Speech* 60, 110–122. doi: 10.1177/0023830916645374

Biau, E., and Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain Lang.* 124, 143–152. doi: 10.1016/j.bandl.2012.10.008

Bradac, J. J. (1990). "Language attitudes and impression formation," in *Handbook of Language and Social Psychology*, eds H. Giles and W. P. Robinson (London: John Wiley), 387–412.

Cheng, L. R. L. (1999). Moving beyond accent: social and cultural realities of living with many tongues. *Top. Lang. Disord.* 19, 1–10. doi: 10.1097/00011363-199908000-00004

Church, R. B., Alibali, M. W., and Kelly, S. D. (eds) (2017). *Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*. Amsterdam: John Benjamins Publishing Company.

Clark, J. M., and Paivio, A. (1991). Dual coding theory and education. *Educ. Psychol. Rev.* 3, 149–210. doi: 10.1007/BF01320076

Dahl, T. I., and Ludvigsen, S. (2014). How I see what you're saying: the role of gestures in native and foreign language listening comprehension. *Modern Lang. J.* 98, 813–833.

DeJesus, J. M., Hwang, H. G., Dautel, J. B., and Kinzler, K. D. (2017). Bilingual children's social preferences hinge on accent. *J. Exp. Child Psychol.* 164, 178–191. doi: 10.1016/j.jecp.2017.07.005

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I, and Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum. Brain Mapp.* 30, 3509–3526. doi: 10.1002/hbm.20774

Drijvers, L., and Özyürek, A. (2017). Visual context enhanced: the joint contribution of iconic gestures and visible speech to degraded speech comprehension. *J. Speech Lang. Hear. Res.* 60, 212–222. doi: 10.1044/2016_JSLHR-H-16-0101

Efron, D. (1941). *Gesture and Environment: A Tentative Study of Some of the Spatio-Temporal and" Linguistic" Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City, Living under Similar as well as Different Environmental Conditions*. New York, NY: King's crown Press.

Ekman, P. (1972). "Universals and cultural differences in facial expressions of emotion," in *Nebraska Symposium on Motivation*, Vol. 19, ed. J. Cole (Lincoln: University of Nebraska Press), 207–283.

Giles, H. (1977). *Language, Ethnicity and Intergroup Relations*. London: Academic Press.

Gluhareva, D., and Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse demanding situations. *Lang. Teach. Res.* 21, 609–631. doi: 10.1177/1362168816651463

Gluszek, A., and Dovidio, J. F. (2010). The way they speak: a social psychological perspective on the stigma of nonnative accents in communication. *Pers. Soc. Psychol. Rev.* 14, 214–237. doi: 10.1177/1088868309359288

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends Cogn. Sci.* 3, 419–429. doi: 10.1016/s1364-6613(99)01397-2

Graham, J. A., and Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures (1). *Int. J. Psychol.* 10, 57–67. doi: 10.1080/00207597508247319

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Mapp.* 30, 3309–3324. doi: 10.1002/hbm.20753

Gregersen, T. S. (2005). Nonverbal cues: clues to the detection of foreign language anxiety. *For. Lang. Ann.* 38, 388–400. doi: 10.1111/j.1944-9720.2005.tb02225.x

Grosjean, F. (2010). *Bilingual*. Cambridge, MA: Harvard University Press.

Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund: Lund University Press.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition. (*Hommage à Adam Kendon*). *IRAL-Int. Rev. Appl. Linguist. Lang. Teach.* 44, 103–124. doi: 10.1515/IRAL.2006.004

Hannah, B., Wang, Y., Jongman, A., Sereno, J. A., Cao, J., and Nie, Y. (2017). Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers. *Front. Psychol.* 8:2051. doi: 10.3389/fpsyg.2017.02051

Hansen, K., Steffens, M. C., Rakiæ, T., and Wiese, H. (2017). When appearance does not match accent: neural correlates of ethnicity-related expectancy violations. *Soc. Cogn. Affect. Neurosci.* 12, 507–515. doi: 10.1093/scan/nsw148

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., and Weber, A. (2012). When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. *J. Cogn. Neurosci.* 24, 878–887. doi: 10.1162/jocn_a_00103

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Hoetjes, M., van Maastricht, L., & van der Heijden, L. (2019). Gestural training benefits L2 phoneme acquisition: findings from a production and perception perspective. in *Proceedings of the Conferene on 6th Gesture and Speech in Interaction*. Paderborn.

Holle, H., Obleser, J., Rueschemeyer, S.-A., and Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage* 49, 875–884. doi: 10.1016/j.neuroimage.2009.08.058

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol. Bull.* 137, 297–315. doi: 10.1037/a0022128

Howie, J. M., and Howie, J. M. (1976). *Acoustical Studies of Mandarin Vowels and Tones*, Vol. 18. Cambridge: Cambridge University Press.

Huang, X., Kim, N., and Christianson, K. (2019). Gesture and vocabulary learning in a second language. *Lang. Learn.* 69, 177–197. doi: 10.1111/lang.12326

Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapratto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565

Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0

Jungheim, N. (2001). "The unspoken element of communicative competence: evaluating language learners' nonverbal behavior," in *A Focus on Language Test Development: Expanding the Language Proficiency Construct across a Variety of Tests*, eds T. Hudson and J. Brown (Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center), 1–35.

Jussim, L., Coleman, L. M., and Lerch, L. (1987). The nature of stereotypes: a comparison and integration of three theories. *J. Pers. Soc. Psychol.* 52, 536–546. doi: 10.1037/0022-3514.52.3.536

Kelly, S. D. (2017). "Exploring the boundaries of gesture-speech integration during language comprehension," in *Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*, eds R. B. Church, M. W. Alibali, and S. D. Kelly (Amsterdam: John Benjamins Publishing), 243–265.

Kelly, S. D., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Lang. Cogn. Process.* 24, 313–334. doi: 10.1080/01690960802365567

Kelly, S. D., Özyürek, A., and Maris, E. (2010). Two sides of the same coin: speech and gesture manually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327

Kendon, A. (1997). Gesture. *Annu. Rev. Anthropol.* 26, 109–128. doi: 10.1146/annurev.anthro.26.1.109

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kendon, A. (2017). Pragmatic functions of gestures: some observations on the history of their study and their nature. *Gesture* 16, 157–175. doi: 10.1075/gest.16.2.01ken

Kinzler, K. D., Corriveau, K. H., and Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Dev. Sci.* 14, 106–111. doi: 10.1111/j.1467-7687.2010.00965.x

Kinzler, K. D., Dupoux, E., and Spelke, E. S. (2007). The native language of social cognition. *PNAS Proc. Natl. Acad. Sci. U. S. A.* 104, 12577–12580. doi: 10.1073/pnas.0705345104

Kinzler, K. D., Shutts, K., DeJesus, J., and Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Soc. Cogn.* 27, 623–634. doi: 10.1521/soco.2009.27.4.623

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Lang. Cogn. Process.* 24, 145–167. doi: 10.1080/01690960802586188

Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005

Kushch, O., Igualada, A., and Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Lang. Cogn. Neurosci.* 33, 992–1004. doi: 10.1080/23273798.2018.1435894

LaSasso, C., Crain, K., and Leybaert, J. (2003). Rhyme generation in deaf students: the effect of exposure to cued speech. *J. Deaf Stud. Deaf Educ.* 8, 250–252. doi: 10.1093/deafed/eng014

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

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about nonnative English speakers in the United States. *J. Socioling.* 7, 348–364. doi: 10.1111/1467-9481.00228

Lippi-Green, R. (2012). *English with an Accent: Language, Ideology, and Discrimination in the United States*, 2nd Edn. London: Routledge.

Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Front. Psychol.* 5:1467. doi: 10.3389/fpsyg.2014.01467

Macedonia, M., and Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind Brain Educ.* 8, 74–88. doi: 10.1111/mbe.12047

Maricchiolo, F., Gnisci, A., Bonaiuto, M., and Ficca, G. (2009). Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Lang. Cogn. Process.* 24, 239–266. doi: 10.1080/01690960802159929

Matsumoto, D., and Hwang, H. S. (2013). Emblematic gestures (Emblems). *Encycl. Cross Cult. Psychol.* 2, 464–466. doi: 10.1002/9781118339893.wbeccp188

McCafferty, S. G., and Stam, G. (eds) (2009). *Gesture: Second Language Acquisition and Classroom Research*. Abingdon: Routledge.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychol. Rev.* 92, 350–371. doi: 10.1037/0033-295X.92.3.350

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2006). *Gesture and Thought*. Chicago, IL: University of Chicago Press.

McNeill, D., Cassell, J., and McCullough, K. E. (1994). Communicative effects of speech-mismatched gesture. *Res. Lang. Soc. Interact.* 27, 223–237. doi: 10.1207/s15327973rlsi2703_4

Morett, L. M. (2014). When hands speak louder than words: the role of gesture in the communication, encoding, and recall of words in a novel second language. *Mod. Lang. J.* 98, 834–853. doi: 10.1111/j.1540-4781.2014.12125.x

Morett, L. M., and Chang, L. Y. (2015). Emphasizing sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Lang. Cogn. Neurosci.* 30, 347–353. doi: 10.1080/23273798.2014.923105

Neu, J. (1990). "Assessing the role of nonverbal communication in the acquisition of communicative competence in L2," in *Developing Communicative Competence in a Second Language*, eds R. C. Scarcella, E. S. Andersen, and S. D. Krashen (New York, NY: Newbury House), 121–138.

Novack, M. A., and Goldin-Meadow, S. (2017). *Understanding Gesture as Representational Action. Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*. Amsterdam: John Benjamins Publishing Company.

Obermeier, C., Dolk, T., and Gunter, T. C. (2011). The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. *Cortex* 48, 857–870. doi: 10.1016/j.cortex.2011.02.007

Özyürek, A. (2017). "Function and processing of gesture in the context of language," in *Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*, eds R. B. Church, M. W. Alibali, and S. D. Kelly (Amsterdam: John Benjamins Publishing).

Paivio, A. (1990). Dual coding theory: retrospect and current status. *Can. J. Psychol.* 45, 255–287. doi: 10.1037/h0084295

Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annu. Rev. Appl. Ling.* 26, 219–233. doi: 10.1017/S0267190506000110

Poggi, I. (2008). Iconicity in different types of gestures. *Gesture* 8, 45–61. doi: 10.1075/gest.8.1.05pog

Popelka, G. R., and Berger, K. W. (1971). Gestures and visual speech reception. *Am. Ann. Deaf* 116, 434–436.

Pourtois, G., de Gelder, B., Bol, A., and Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex* 41, 49–59. doi: 10.1016/S0010-9452(08)70177-1

Pouw, W., Paxton, A., Harrison, S. J., and Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proc. Natl. Acad. Sci. U.S.A.* 117, 11364–11367. doi: 10.1073/pnas.2004163117

Poyatos, F. (1983). *New Perspectives in Nonverbal Communication. Studies in Cultural Anthropology, Social Psychology, Linguistics, Literature and Semiotics*. Oxford: Pergamon Press.

Rubin, D. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Res. High. Educ.* 33, 511–531. doi: 10.1007/bf00973770

Schneller, R. (1988). "The Israeli experience of crosscultural misunderstandings: insights and lessons," in *Cross-Cultural Perspectives in Nonverbal Communication*, ed. F. Poyatos (Toronto, ON: Hogrefe), 153–173.

Sime, D. (2006). What do learners make of teachers' gestures in the language classroom? *Int. Rev. Appl. Ling. Lang. Teach.* 44, 211–230. doi: 10.1515/IRAL.2006.009

Skipper, J. I. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 369:20130297. doi: 10.1098/rstb.2013.0297

So, W. C., Sim, C. C. H., and Low, W. S. J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: is meaning in gesture important for memory recall?. *Lang. Cogn. Process.* 27, 665–681. doi: 10.1080/01690965.2011.573220

Sueyoshi, A., and Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Lang. Learn.* 55, 661–699. doi: 10.1111/j.0023-8333.2005.00320.x

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054

Van den Stock, J., Righart, R., and De Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion* 7:487. doi: 10.3389/fnhum.2014.00053

Vance, T. J. (2008). *The Sounds of Japanese*. Cambridge: Cambridge University Press.

Wang, L., and Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia* 51, 2847–2855. doi: 10.1016/j.neuropsychologia.2013.09.027

Willems, R. M., Özyürek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Wu, Y. C., and Coulson, S. (2007). How iconic gestures enhance communication: an ERP study. *Brain Lang.* 101, 234–245. doi: 10.1016/j.bandl.2006.12.003

Zheng, A., Hirata, Y., and Kelly, S. D. (2018). Exploring the effects of imitating hand gestures and head nods on L1 and L2 Mandarin tone production. *J. Speech Lang. Hear. Res.* 61, 2179–2195. doi: 10.1044/2018_jslhr-s-17-0481

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 handling editor declared a past collaboration with one of the authors SK.

ORIGINAL RESEARCH
published: 23 September 2020
doi: 10.3389/fpsyg.2020.575628

Teaching Analogical Reasoning With Co-speech Gesture Shows Children Where to Look, but Only Boosts Learning for Some

Katharine F. Guarino and Elizabeth M. Wakefield*

Department of Psychology, Loyola University Chicago, Chicago, IL, United States

In general, we know that gesture accompanying spoken instruction can help children learn. The present study was conducted to better understand *how* gesture can support children's comprehension of spoken instruction and whether the benefit of teaching though speech and gesture over spoken instruction alone depends on differences in cognitive profile – prior knowledge children have that is related to a to-be-learned concept. To answer this question, we explored the impact of gesture instruction on children's analogical reasoning ability. Children between the ages of 4 and 11 years solved scene analogy problems before and after speech alone or speech and gesture instruction while their visual attention was monitored. Our behavioral results suggest a marginal benefit of gesture instruction over speech alone, but only 5-year-old children showed a distinct advantage from speech + gesture instruction when solving the post-instruction trial, suggesting that at this age, children have the cognitive profile in place to utilize the added support of gesture. Furthermore, while speech + gesture instruction facilitated effective visual attention during instruction, directing attention away from featural matches and toward relational information was pivotal for younger children's success post instruction. We consider how these results contribute to the gesture-for-learning literature and consider how the nuanced impact of gesture is informative for educators teaching tasks of analogy in the classroom.

Keywords: gesture, learning, visual attention, eye tracking, analogical reasoning

OPEN ACCESS

Edited by:
Kazuki Sekine,
Waseda University, Japan

Reviewed by:
Ercenur Ünal,
Özyeğin University, Turkey
Fred Paas,
Erasmus University Rotterdam,
Netherlands

****Correspondence:***
Katharine F. Guarino
kguarino@luc.edu

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

Received: *23 June 2020*
Accepted: *20 August 2020*
Published: *23 September 2020*

Citation:
Guarino KF and Wakefield EM (2020) Teaching Analogical Reasoning With Co-speech Gesture Shows Children Where to Look, but Only Boosts Learning for Some. Front. Psychol. 11:575628. doi: 10.3389/fpsyg.2020.575628

INTRODUCTION

Gestures – movements of the hands that are naturally used in conversation and express ideas through their form and movement trajectory – help children learn. This has been found across domains, including mathematics (Singer and Goldin-Meadow, 2005; Cook et al., 2013), symmetry (Valenzeno et al., 2003), conservation (Church et al., 2004; Ping and Goldin-Meadow, 2008), and word learning (Wakefield et al., 2018a). And while this function of gesture is well-established, the mechanism by which gesture supports children's learning, and how individual differences between children impacts the effectiveness of incorporating gesture into instruction, are not fully understood.

One way gesture is thought to help children learn is by grounding and disambiguating the meaning of spoken instruction (e.g., Alibali and Nathan, 2007; Ping and Goldin-Meadow, 2008).

When learning a new concept, children may struggle to understand the meaning of spoken instruction and fail to see connections between a teacher's speech and their use of supportive materials like equations, figures, or diagrams. Gestures facilitate connections between spoken language and these physical supports by directing attention to key components of a problem being taught or providing a visual depiction of an abstract concept through hand shape or movement trajectory (e.g., McNeill, 1992; Altmann and Kamide, 1999; Huettig et al., 2011; Wakefield et al., 2018b). For example, when being taught the concept of mathematical equivalence – the idea that two sides of an equation are equal to one another (e.g., $2 + 5 + 8 = \_ + 8$) – eye tracking results show that children follow along with spoken instruction more effectively if it is accompanied by gesture than if the concept is explained through speech alone. Importantly, children's ability to follow along with spoken instruction when it was accompanied by gesture predicted their ability to correctly solve mathematical equivalence problems beyond instruction (Wakefield et al., 2018b).

However, incorporating gesture may not support all children's understanding of spoken instruction to the same extent. Although prior work suggests that gesture supports children's learning, there are nuances to when gesture is beneficial: Children's pre-existing knowledge related to a domain – which we will refer to as their *cognitive profile* – can impact whether they learn from gesture instruction. For example, Wakefield and James (2015) taught children the concept of a palindrome (i.e., a word that reads the same forward and backward) through speech-alone or speech + gesture instruction. They considered whether the impact of gesture was affected by children's relevant cognitive profile – in this case, their phonological ability, as the task relied heavily on understanding how sounds in words fit together. Children with high phonological ability benefitted more from speech + gesture instruction than speech-alone instruction, but children with low phonological ability did not show this advantage, suggesting that children need some degree of pre-existing knowledge within the domain to utilize gesture. In this case, the authors argued that gesture could not clarify spoken instruction unless children had a certain level of phonological awareness.

Although not considered by Wakefield and James, there may also be a developmental point when children are on the brink of understanding a concept and have a sufficiently developed cognitive profile that they need just a small boost from instruction to master a concept. In this case, incorporating gesture into instruction might not be any more powerful than spoken instruction alone. There may be a 'sweet spot' where children have enough foundational knowledge and cognitive abilities related to a concept that gesture can clarify spoken instruction and boost their learning, while children far below or above this developmental point do not show an advantage when learning through gesture.

In the present study, to better understand how gesture can support children's understanding of spoken instruction and whether the benefit of teaching though speech and gesture over spoken instruction alone depends on differences in cognitive profile, we explore the impact of gesture in analogical reasoning. Analogical reasoning is the ability to identify underlying schematic or relational structure shared between representations. In its mature form, it is a powerful cognitive mechanism that contributes to a range of skills unique to humans (for review see Gentner and Smith, 2013). For the purpose of the present study, analogical reasoning is a useful testbed because it is a domain that requires disambiguating complex verbal information, and because the relevant cognitive profile for solving analogies shows protracted development across early childhood (e.g., Richland et al., 2006; Thibaut et al., 2010; Thibaut and French, 2016; Starr et al., 2018).

One of the predominant types of analogy task used to assess the development of children's analogical reasoning ability are scene analogies, in which children are asked to examine two scenes (e.g., a source and target scene) which contain both relational similarities and featural similarities. When prompted to solve a scene analogy, children are asked to identify an item in the target scene that corresponds *relationally* to a prompted item in the source scene. However, children often choose an item that corresponds *featurally* to the prompted item instead of the relationally similar item. This type of 'featural match' is one item in a target scene that is not incorporated in the relation of focus, but has great surface similarity to the prompted item in a source scene (Richland et al., 2006). For example, a source scene might show a boy chasing a girl (relation of *chasing*), with the boy prompted. The corresponding target scene would contain a dog chasing a cat (relation of *chasing*) and a second boy (the featural match). Here, the dog would be the correct relational choice, and the second boy would be the incorrect featural match. Young children find it difficult to disengage from the featural match (i.e., another boy that is similar in appearance to the prompted boy) in favor of a relational match (i.e., the other thing that is chasing). This focus on surface features, or perceptual similarities, rather than relational information is a common pitfall for children (Gentner, 1988) that they may not fully overcome until they are 9–11 years of age (Richland et al., 2006).

Because incorporating gesture in instruction can direct children's visual attention effectively to key components of a problem in other domains, such as mathematics instruction (Wakefield et al., 2018b), gesture should also be able to facilitate effective visual attention in problems of analogy. Gesture should be able to clearly indicate, and disambiguate, which items a teacher is referring to when providing spoken instruction, so that children are focused on items and relations relevant for successful solving and do not attend to irrelevant items. When considering the previous example of a scene analogy, a teacher is likely to align the important relations through speech, stating that the boy is chasing the girl, and the dog is chasing the cat. In theory, this type of statement, which highlights structural similarities between contexts, should orient children's attention to the items involved in the relation of chasing, and, thereby, facilitate an analogical comparison (e.g., Gentner, 1983, 2010; Markman and Gentner, 1993; Namy and Gentner, 2002). However, when a featural match is present, this spoken instruction by itself may leave some ambiguity in terms of *which* boy is being discussed. Children may focus their attention on one or both boys (i.e., the boy in the chasing relation and the featural match) and miss the important connections being drawn between the relations in the source and target scenes. Indeed, we know from eye tracking studies that children who incorrectly solve analogical reasoning problems tend to focus their visual attention on the featural match, and ignore relational information (Thibaut and French, 2016; Glady et al., 2017; Starr et al., 2018; Guarino et al., 2019). Instruction that incorporates gesture may help young children understand which boy is relevant to the task and direct their attention away from irrelevant featural matches.

But will gesture instruction provide the same boost to all children who struggle to solve analogical reasoning problems? The determining factor may be a child's cognitive profile relevant to analogical reasoning ability, comprised of effective inhibitory control and working memory. Inhibitory control allows an individual to inhibit more salient, featural match responses, and select a less salient, but correct, relational match (e.g., Viskontas et al., 2004; Richland et al., 2006). Working memory allows an individual to simultaneously process multiple contexts and pieces of information present in an analogy (e.g., Gick and Holyoak, 1980; Halford, 1993; Simms et al., 2018). Due to the protracted development of these cognitive capacities, analogical reasoning similarly develops gradually over time, with initial stages presenting in children as young as 3–5 years old and maturing into adolescence (e.g., Alexander et al., 1987; Goswami and Brown, 1989; Rattermann and Gentner, 1998). In the case of a scene analogy, Richland et al. (2006) find that children have difficulty ignoring featural matches in favor of relational matches until they are 9–11-years-old, with children showing an increase in successful problem solving between the ages of 3 and 11, as children's cognitive profiles develop.

With this protracted development of cognitive profile in mind, we might expect differences in the effectiveness of gesture instruction. For very young children their inhibitory control and working memory may be so limited that they may not be able to capitalize on gesture's ability to index spoken instruction to referents in a scene analogy, and therefore, gesture may not be helpful for disambiguating complex verbal instruction. However, for slightly older children, we may find that gesture provides the exact boost they need: They may have the cognitive profile in place to benefit from instruction, and gesture may give them an extra boost by literally pointing them in the right direction to help them make sense of spoken instruction. For even older children with high inhibitory control and working memory capacity, who typically demonstrate near-adult like ability on problems of analogy, receiving spoken instruction, even without gesture, may be enough support for understanding the structure of analogies.

Present Study

We test these predictions in the present study. To do this, we compare how children across a wide age range (4–11-year-olds) solve scene analogy problems before or after speech alone or speech and gesture instruction while monitoring their visual attention with eye tracking. Using a wide age range will allow us to understand how cognitive profile contributes to the effectiveness of gesture instruction. Using eye tracking will allow us to understand how gesture aids in disambiguation of spoken instruction meant to refer to an item within a relation, that could instead be linked to a featural match. Through this approach, we will address three questions: (1) Do children benefit differently from speech alone versus speech and gesture instruction on analogical reasoning based on their age (as a proxy for cognitive profile)? (2) Can we find evidence that gesture instruction helps disambiguate spoken instruction, and does this depend on age? (3) Do looking patterns associated with type of instruction impact whether children at different ages learn from instruction? Results will add to our general understanding of the mechanisms by which children learn and explore the nuances of when gesture may or may not help beyond spoken instruction. And by focusing on analogical reasoning, we also explore the utility of gesture instruction in a domain that is important for academic success and has been understudied in the gesture-for-learning literature.

MATERIALS AND METHODS

Participants

Children between the ages of 4 and 11 years old (N = 323; 159 females) participated in the present study during a visit to a science museum[1]. Children were randomly assigned to one of two conditions ($n_{speech\text{-}alone}$ = 160; $n_{speech+gesture}$ = 163), with a target of ∼20 children per age group in each condition. An additional 62 children participated in the study but were excluded from analyses for eye tracker malfunction (n = 20), parental involvement (n = 7), language barrier (n = 2), lack of response from participant (n = 7), poor eye tracking (n = 3), and experimenter error (n = 23). Two participants decided they did not want to continue before being assigned a condition. Informed consent was obtained from a parent or guardian of each participant, and verbal assent was obtained from children. Children participated individually in one 3–5 min experimental session and received stickers as compensation.

Materials

Warm-Up Examples

Children were shown two scenes depicting relations occurring between items (see Appendix A for items). For example, a scene showed one animal (e.g., elephant) reading to another animal (e.g., rabbit), and another scene showed an animal (e.g., duck) on top of another animal (e.g., cow). Instruction was provided that highlighted the relation of interest (i.e., *patterns* of 'reading' and 'on top of'). These trials served to familiarize children with our use of the term *pattern* and how items can be *relationally associated*.

[1] Although we did not collect demographic information from individuals, our sample was representative of the general profile of museum visitors. According to museum reports based on short surveys with museum visitors, visitors to the museum represent a number of different racial and ethnic backgrounds (70% White, 10% Hispanic, 6% African American, 6% Asian, 5% Other, <1% Native American, Native Hawaiian), and are also diverse in socioeconomic status, based on self- report measures of perceived socioeconomic status (13% lower or lower-middle class, 54% middle class, 33% upper middle or upper class) and parent or guardian's highest level of formal education (1% < high school diploma, 18% high school diploma, 16% associates degree, 35% bachelor's degree, 21% master's degree, 7% Ph.D., or other terminal professional degree, 3% not reporting).

Pre- and Post-instruction Stimuli

Two scene analogy problems (see Appendix A for items) were selected from a data set created by Guarino et al. (2019), that were based on the structure used by Richland et al. (2006). Scene analogies have been used in a number of other studies assessing the development of children's analogical reasoning ability (e.g., Morrison et al., 2004; Richland et al., 2006, 2010; Gordon and Moser, 2007; Krawczyk et al., 2010; Morsanyi and Holyoak, 2010; Glady et al., 2016; Simms et al., 2018). Previous work has found that children as young as 3–4 years old can successfully solve scene analogy problems when there is not a featural match present just over half of the time (Richland et al., 2006). And by age 9–11 children are fairly proficient at solving scene analogies, even when featural matches are present (Richland et al., 2006). Therefore, this analogy format is particularly useful for assessing analogical reasoning ability across the age range utilized in the present study because it encompasses the entire developmental trajectory of this ability.

Each problem included a pair of scenes, a source scene on the left, and a target scene on the right. Scenes depicted the relation *chasing* occurring between items (i.e., animals or people; **Figure 1**). Source scenes contained five items: the two items within the relation of chasing, and three additional items (i.e., neutral inanimate objects that were not involved in the relation of chasing). One of the items within the source scene relation was circled. Target scenes also contained five items: the two items within the relation, two additional items, and a featural match. The *featural match* was similar to the circled source-scene item and centrally located, increasing the likelihood that the item would draw participants' attention.

Figure 1 shows an example of a chasing *source* and *target* scene. The source scene on the left shows a tiger chasing a woman (items within the *chasing* relation), and a dog-house, jeep and plant (neutral items). The corresponding target scene on the right shows a lion chasing a horse (items within the *chasing* relation), a barn and soccer ball (neutral items), and a tiger (a featural match item that is superficially similar to the prompted tiger in the source scene).

The directionality of relations within a pair of scenes was reversed to avoid children making choices based on spatial location alone. For example, in **Figure 1**, the direction of chasing is right to left in the source scene (the tiger on the right is chasing the woman on the left), whereas the direction of chasing is left to right in the target scene (the lion on the left is chasing the horse on the right). Children were presented with printed copies of scene analogies. Stimuli were bound in a binder, with one pair of scenes presented at a time.

Instruction Stimuli

Similar to pre- and post-instruction trials, printed instruction stimuli included two scenes in which a chasing relation was depicted in both scenes, and a featural match was located in the target scene (see Appendix A for items). Unlike pre- and post-instruction trials, no item was circled in the instruction stimuli.

Eye Tracker

Eye tracking data were collected via corneal reflection using a Tobii Pro Glasses 2. Tobii software was used to perform a 1-point calibration procedure. This step was followed by the collection and integration of gaze data using Tobii Pro Lab (Tobii Technology, Sweden). Data were extracted on the level of individual fixations as defined by Tobii Pro Lab software—an algorithm that determined if two points of gaze data are within a preset minimum distance from one another for a minimum of 100 ms, allowing for the exclusion of eye position information during saccades. After extraction, fixations were manually mapped by research assistants. Individual fixations were classified as either oriented toward one of the items of interest within the scenes (e.g., to the item chasing in the source scene, to the item being chased in the source scene, to the featural match, etc.), other areas around the items within the scenes, or the space surrounding the scenes. Research assistants assigned each fixation to an area of interest (AOI), based on its location (e.g., if a fixation was located on or within the immediate area surrounding the featural match, it was manually mapped as a featural match fixation). For more details about manual mapping, see Appendix B.

Procedure

Children participated individually at a table in a corner of the museum floor. Children were told they were going to play a picture game while wearing eye tracking glasses. After a brief explanation that the purpose of the glasses is to 'help us see what you see,' an experimenter fitted them with the glasses. Children were seated approximately 40 cm in front of the printed stimuli next to an experimenter. The printed stimuli were displayed in a binder mounted on an easel. This allowed the experimenter to quickly flip between stimuli and gesture to the stimuli during instruction trials if a child was assigned to the speech + gesture condition. It also ensured proper eye tracking – children could see the stimuli directly in front of them, and did not have to look down toward the table, which would have disrupted our ability to capture their visual attention via the eye glasses. Children's position was calibrated and adjusted if necessary, and they were asked to remain as still as possible during the rest of the game while eye tracking data were collected.

First, the experimenter explained the relational pattern in the two warm-up trials, meant to help promote relational thinking (see section "Materials and Methods" for details). Next, children completed one pre-instruction trial. After orienting children to the two scenes presented simultaneously (e.g., one side has blue edges and one has green edges), children were asked, "*Which thing in the picture with the blue edges is in the same part of the pattern as the circled thing in the picture with the green edges?*" An item in the source scene (e.g., in **Figure 1**, green edges) was circled and had a corresponding relational item and featural match in the target scene (e.g., in **Figure 1**, blue edges). All stimuli used in this task can be seen in Appendix A and additional details about the stimuli can be found in the "Materials and Methods" section. The task was self-paced, but if no response was given after a few seconds, the children were re-prompted by the experimenter.

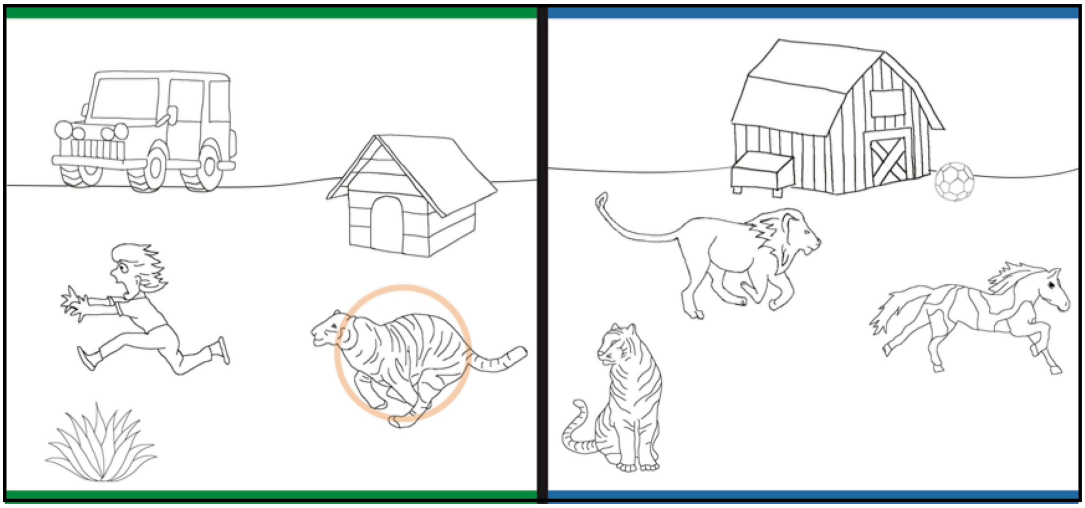


FIGURE 1 | Example trial of a chasing relation stimulus.

Following the pre-instruction trial children were asked to pay attention to two instruction trials to learn about the pattern in the pictures. Children were randomly assigned to receive *speech-alone* instruction or *speech + gesture* instruction provided by the experimenter. In her instruction, the experimenter described chasing relations and similarities between items from a source and target scene, displayed in front of the child. For example, in a scene analogy problem with a boy chasing a girl in a source scene and a dog chasing a cat in a target scene with a featurally matched boy present, the experimenter said, "*See, the boy is chasing the girl, and the dog is chasing the cat. This means the boy is in the same part of the pattern as the dog because they are both chasing, and the girl is in the same part of the pattern as the cat because they are both being chased.*" The ambiguity of this instruction occurs when the boy in the source scene is referenced, because there is also a featurally similar boy in the target scene (i.e., the featural match). When the boy is mentioned in speech it may be difficult for children to reconcile *which* boy is being discussed: the one in the relation of chasing or the featural match. This confusion or ambiguity could contribute to difficulty identifying the relational structure in an analogy problem.

In the speech + gesture condition, the experimenter provided the same spoken instruction, accompanied by gestures that emphasized items and relations. In the example above, when the experimenter said '*The boy is chasing the girl*,' a sweeping movement of the index finger traced a path from the boy to the girl, highlighting the chasing relation. The same sweeping gesture was used when the experimenter said '*. . . and the dog was chasing the cat.*' Then, deictic gestures – pointing gestures used to indicate objects or locations – were used to simultaneously reference the items that were in the same parts of the relations. Items were indicated by a pointed index finger on each hand. When the experimenter said, '*This means the boy is in the same part of the pattern as the dog because they are both chasing,*' simultaneous deictic gestures pointed to the boy and the dog. Similarly, when the experimenter said, '*. . .and the girl is in the same part of the pattern as the cat because they are both being chased,*' simultaneous deictic gestures pointed to the girl and the cat (see **Figure 2** for an example of children's view during training).

Finally, a post-instruction trial was administered after children viewed the instructional trials, with an identical prompt and procedure as used during the pre-instruction trial.

Measures of Visual Attention

Measure of Attention During Pre- and Post-instruction Trials

Visual attention during pre- and post-instruction trials was quantified by generating areas of interest that represent different portions of the participant's field of view using Tobii Pro Lab. There were 11 AOIs in total (see Appendix B). The AOIs encompassed regions within the scene pairs and areas in the field of view that were outside of the scene analogy. This included an AOI for each of the items in the scenes (items in chasing relations, featural match, and neutral items), AOIs for when the participant fixated on the experimenter, on the experimenter's gesture, and on their own hands, and an AOI for looking elsewhere in the museum. Proportion of time spent looking to each AOI was then calculated by dividing the time looking to an AOI during a trial by the total time looking during a trial. For the sake of the present analyses, we focused on the AOI representing the featural match. Children's ability to avoid featural matches is one of the key issues children overcome as they develop successful analogical reasoning. By assessing visual attention to the featural match we can assess whether gesture is more effective than speech alone for driving attention away from irrelevant featural components.

Measures of Attention During Instruction

Attention during instruction was quantified in two ways: (1) children's ability to synchronize their visual attention with spoken instruction and (2) 'check-ins' with the featural match during ambiguous spoken instruction.

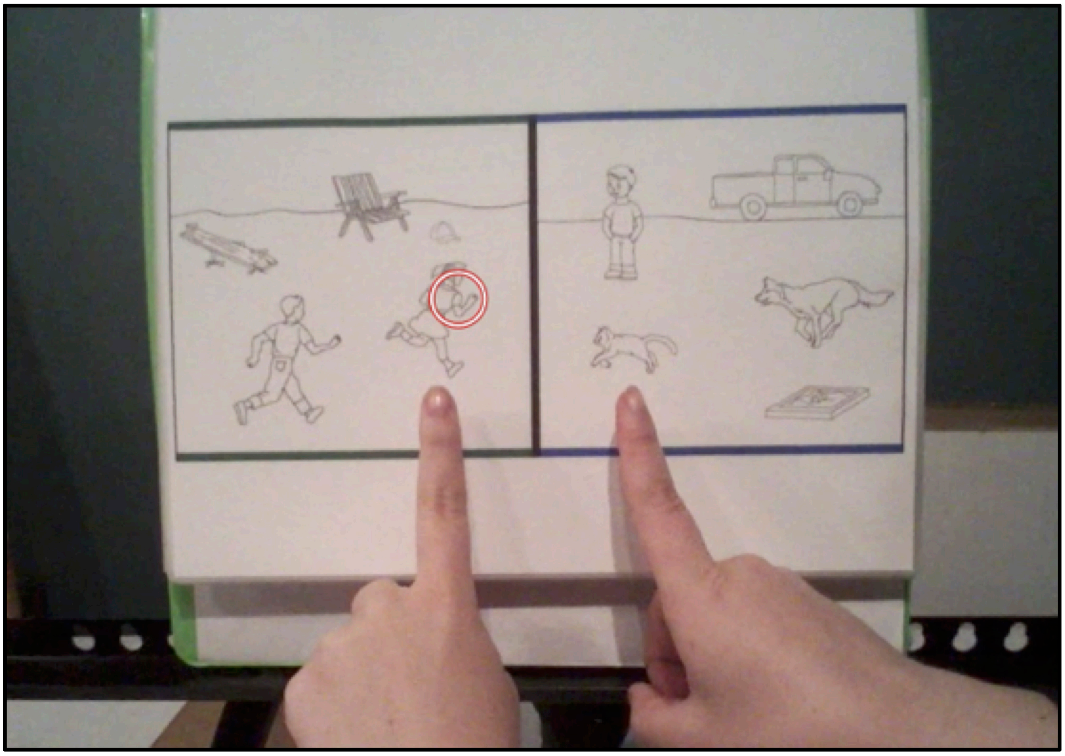


FIGURE 2 | Example of children's view during a speech + gesture training trial. The red circle shows where the child was focusing his or her visual attention at this moment of instruction.

Following score

Because previous work suggests that gesture can help children follow along with spoken instruction and that this is predictive of learning (Wakefield et al., 2018b), we calculated a 'following score' for each instruction trial. Following scores were calculated by creating four time segments in which different relational comparisons were made by the experimenter and assessing whether children looked to AOIs highlighted in speech during each segment (i.e., during a given segment, children received a score of '1' if they looked to the relevant AOIs as they were labeled in speech and a '0' if they did not). Children could receive a score of 0 to 4 on each training trial, and scores were averaged across the two training trials to generate an overall following score for each child. The average following score was used in analyses.

Check-in score

Check-ins with the featural match are instances when the item that is perceptually similar to the featural match is referenced in speech and simultaneously fixated on by the child. In each instruction trial, there were two time segments during which a check-in could occur. For example, in the instruction trial depicting a boy chasing a girl in the source scene and a featural match boy in the target scene, the two relevant time segments occur when the experimenter said '*The boy is chasing the girl*' and '*The boy is in the same part of the pattern as the dog because they are both chasing.*' For each segment, a child would receive a score of 1 if they looked to the featural match boy in the target scene rather than the boy in the source scene. Children would receive a score of 2 for a given instruction trial if they looked at the featural match boy during both time segments in which the boy in the relation was mentioned. Thus, whereas a score of 4 is possible for following score, a score of 2 is possible for check-in score. Check-in scores from the two instruction trials were averaged to generate an overall check-in score for each child. The average 'check in' score was used in analyses.

RESULTS

All analyses were conducted using R Studio (version 1.1.456), supported by R version 3.6.0. Analyses relied on the *stats* package, which allows for ANOVA and regression modeling (R Core Team, 2017). When running binomial generalized linear regression models assessing the impact of condition on accuracy or choice of the featural match at pre- and post-instruction, the speech-alone condition was set as the baseline condition and compared against the speech + gesture condition. For analyses of visual attention, which did not use a binomial outcome, generalized linear regression models were used. Again, speech-alone was set as the reference level for these analyses.

Before addressing our main questions of interest, we wanted to establish (1) that there were no significant performance differences pre-instruction between children who had been randomly assigned to the speech-alone versus speech + gesture condition – we found that there were not: Both across all children and within age groups, there were no condition differences between pre-instruction accuracy or choice of the featural match (all *ps* > 0.1), and (2) that age could serve as a proxy for cognitive profile. To do this, we asked whether children's ability to solve analogical reasoning problems could be predicted by age and visual attention before instruction. We reasoned that previous work has shown that as children's inhibitory control and working memory improve, they are more likely to succeed on analogical reasoning problems (e.g., Doumas et al., 2018; Simms et al., 2018), and that children with lower inhibitory control look more to the featural match when solving scene analogy problems (Guarino et al., under revision), thus, finding that age was predictive of these measures would suggest that age can serve as a proxy for cognitive profile.

While only 20% of children correctly answered the pre-instruction trial, there was a main effect of age when predicting accuracy, such that older children were more likely to answer the problem correctly than younger children (**Figure 3**, β = 0.18, *SE* = 0.06, *t* = 2.89, *p* = 0.004), replicating previous work (e.g., Richland et al., 2006). And, as with previous studies using scene analogy problems, we found that the most common error children made was to choose the featural match – 64% of children made this type of error. In terms of visual attention, we assessed whether children's proportion looking to the featural match before instruction predicted their performance, as this is a key looking pattern associated with making featural errors (e.g., Thibaut et al., 2010; Thibaut and French, 2016; Guarino et al., 2019). On average, children who correctly answered the pre-instruction trial allocated less of their attention to the featural match (*M* = 0.12, *SD* = 0.08) than children who made featural errors (*M* = 0.14, *SD* = 0.08). Models predicting accuracy by visual attention to the featural match showed that proportion looking to the featural match was negatively related with accuracy (β = −0.00, *SE* = 0.00, *t* = −2.22, *p* = 0.026) and positively related with featural errors (β = 0.00, *SE* = 0.00, *t* = 3.51, *p* < 0.001). In sum, these results replicate previous work finding that prior to instruction, children who are older and attend less to the featural match more successfully solve a scene analogy problem, and provide support for considering age as a proxy for cognitive profile.

Impact of Age and Instruction on Children's Analogical Reasoning Ability

To understand how speech-alone versus speech + gesture instruction affected children's performance on the post-instruction trial, we limited the remainder of our analyses to children who incorrectly answered the pre-instruction trial (speech-alone: *n* = 124; speech + gesture: *n* = 133) – importantly, a similar number of children were excluded from both experimental groups. Our first main question was whether the impact of gesture instruction on children's analogical reasoning is dependent on their cognitive profile (measured by age). Overall, more children in the speech + gesture condition correctly answered the post-instruction trial than children in the speech-alone condition (speech + gesture: 63% vs. speech-alone: 59%). But, from **Figure 4** it is clear that performance is also dependent on age, and when considering performance binned by age, we see that the difference between conditions appears most pronounced for 5-year-olds. To determine whether these patterns were statistically significant, we constructed a generalized linear model with accuracy (0, 1) as the dependent measure, and age, condition (speech-alone, speech + gesture), and an interaction

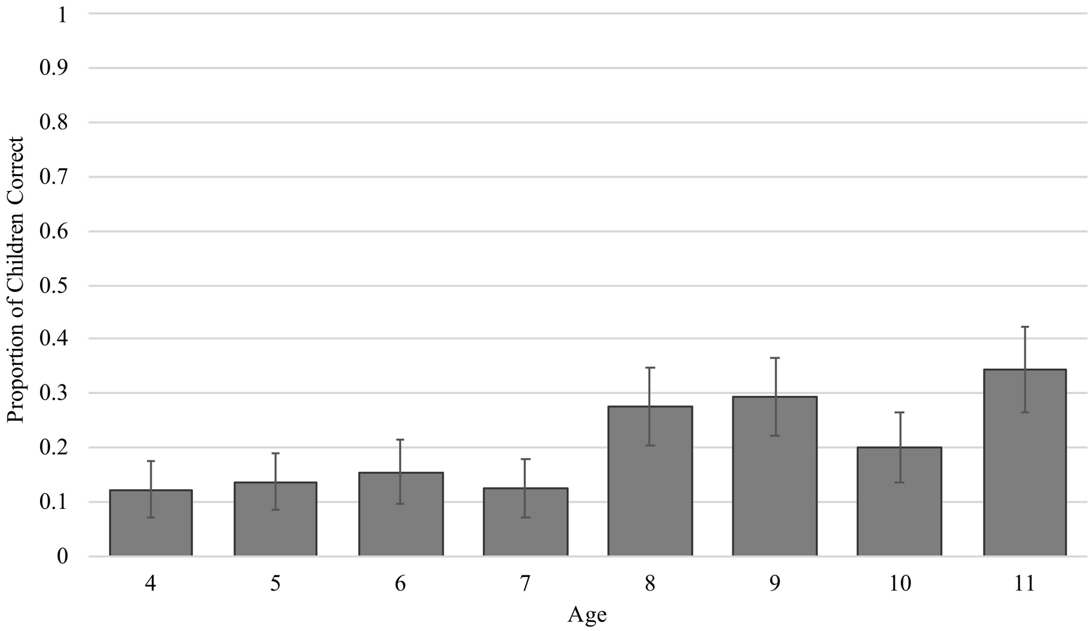


FIGURE 3 | Proportion of children within each age correct on the pre-instruction trial.

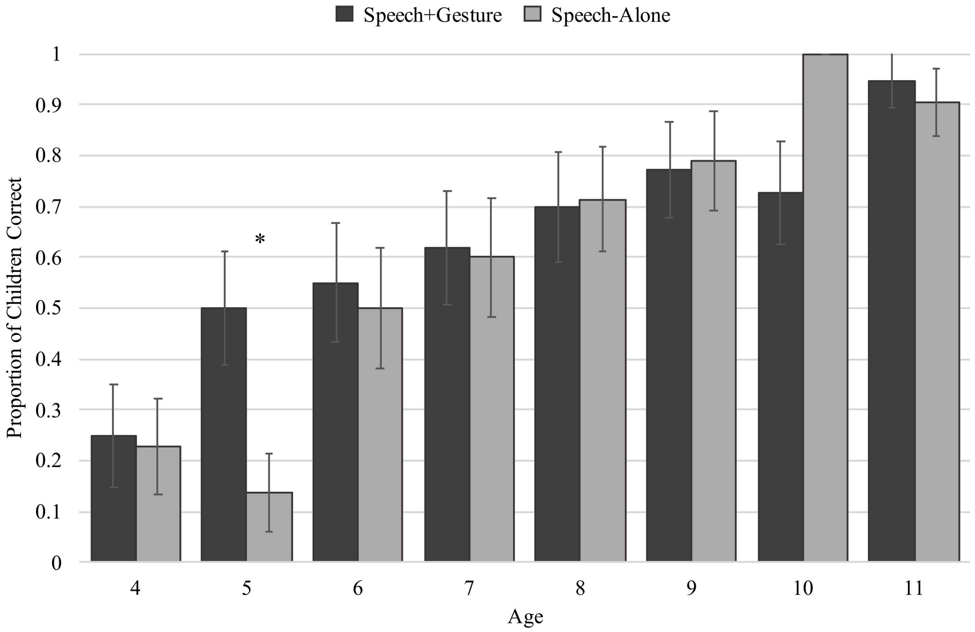


FIGURE 4 | Proportion of children within each age correct on the post-instructional trial separated by condition. *indicates significance at $p > 0.05$

between age and condition as predictors of interest. In line with **Figure 4**, the model revealed a main effect of age, suggesting that older children performed better after instruction than younger children (β = 0.62, *SE* = 0.12, $t = 5.35$, $p < 0.001$), and a trending main effect of condition, suggesting that children improved marginally more after speech + gesture instruction than speech-alone instruction (β = 1.82, *SE* = 1.06, $t = 1.72$, $p = 0.085$). However, these results should be considered within the context of a marginal interaction between age and condition (β = −0.25, *SE* = 0.15, $t = -1.69$, $p = 0.092$), where *post hoc* analyses indicate that only 5-year-old children demonstrate a benefit for speech + gesture compared to speech-alone (β = 1.75, *SE* = 0.89, $t = 1.97$, $p = 0.048$), and for all other children, there was not an effect of condition (*p*s > 0.1). Although this interaction was only marginally significant, this is likely due to the consideration of such a wide age range, with most age groups showing a clear lack of difference in response to instruction condition. The presence of an interaction aligns with the *a priori* hypothesis that gesture may only boost learning beyond speech-alone instruction at certain ages. Given previous work within the analogical reasoning literature that shows 5-year-olds demonstrate greater difficulty with problems incorporating featural matches than older children (e.g., Richland et al., 2006; Simms et al., 2018), it makes sense that gesture would provide these children the most benefit.

Gesture's Effect on Visual Attention During Instruction

Gesture instruction has previously been shown to help children follow along with spoken instruction and facilitate performance on subsequent assessments (Wakefield et al., 2018b). To understand how visual attention might play a role in the marginal behavioral effects of gesture on children's post-instruction performance, we next asked how condition and age influenced children's visual attention during instruction. Here, we used two measures of visual attention: following score and featural match check-in score. Children's following score is an index of whether they looked at relevant referents of the problem (i.e., items involved in the relation of chasing) when the referents were mentioned in spoken instruction. Children's featural match check-in score is an index of whether children attended to the featural match when the instructor's speech was meant to reference an item within a chasing relation, but was ambiguous. Without understanding the context of the analogy, children could associate the spoken referent with either an item within a relevant chasing relation (the item the instructor meant to reference) *or* the featural match to that item (an item that is irrelevant to the analogy). Attending to the featural match may disrupt a child's ability to effectively learn from instruction because it detracts from children's ability to process how the items within the two chasing relations are aligned.

On average, children followed along more successfully with spoken instruction if they were taught through speech + gesture (*M* = 3.08 out of a possible score of 4, *SD* = 1.10) than through speech alone (*M* = 2.20, *SD* = 1.06). **Figure 5** shows following score separated by age and condition and suggests that gesture supports effective following along with instruction for all children. Using a generalized linear model with following score as the dependent measure and age, condition (speech-alone and speech + gesture), and an interaction between age and condition as the predictors of interest, we found a main effect of condition, confirming that speech + gesture instruction supported more effective following than speech-alone instruction

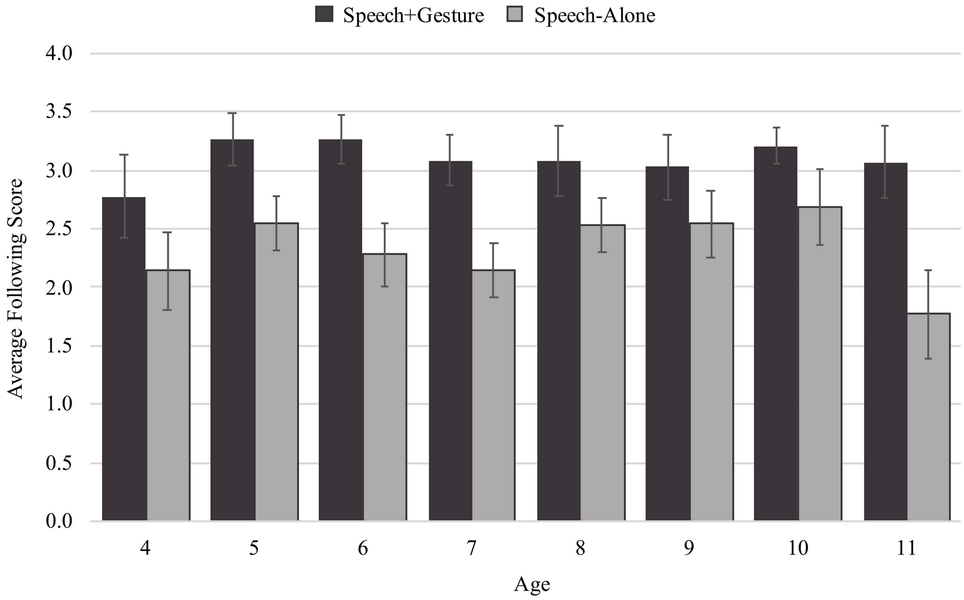


FIGURE 5 | Average following scores split by age and condition.

(β = 1.51, *SE* = 0.26, *t* = 5.81, *p* < 0.001). We also found no main effect of age (β = 0.02, *SE* = 0.06, *t* = 0.28, *p* = 0.783) and no interaction between age and condition (β = 0.12, *SE* = 0.12, *t* = 0.25, *p* = 0.806) suggesting that gesture is a cue that can organize visual attention regardless of a child's age.

Our second measure of visual attention during instruction was how children attended to the featural match, the key component of an analogy that draws children's attention away from the relational information (e.g., Thibaut et al., 2010; Thibaut and French, 2016; Guarino et al., 2019). Specifically, we asked whether children attended to the featural match during the time intervals when the spoken instruction was ambiguous as to whether the instructor was referring to an item within a relation, or the featural match to that item outside of the relation (e.g., Which boy is being referred to: the boy in the relation of chasing or the featural match boy?). Because the lure of featural matches are at the root of young chilldren's difficulties with problems of analogy, the most ambiguous portion of the instruction is when the item that is involved in the relation of chasing and perceptually similar to the featural match is discussed in speech.

On average, children checked-in more with the featural match if they were in the speech-alone condition (*M* = 1.29 out of a possible score of 2, *SD* = 0.48) than in the speech + gesture condition (*M* = 0.71, *SD* = 0.57). But again, the amount of difference between conditions seems to differ by age (**Figure 6**). Using a generalized linear model with check-in score as the dependent measure, and age, condition (speech-alone and speech + gesture), and an interaction between age and condition as the predictors, we found a main effect of condition, such that speech + gesture instruction facilitates fewer check-ins than speech-alone instruction (β = −1.77, *SE* = 0.47, *t* = −3.76, *p* < 0.001), and a main effect of age, such that older children check-in with the featural match less than younger children regardless of the type of instruction received (β = −0.11, *SE* = 0.05, *t* = −2.36, *p* = 0.019). These effects should be interpreted within the context of a trending interaction between condition and age (β = 0.00, *SE* = 0.06, *t* = 1.95, *p* = 0.052). *Post hoc* analyses indicate that this trending interaction results from a developmental shift between younger and older children (**Table 1**): Generally, older children are less likely to show a significant difference in check-in score across conditions, suggesting that they can make use of either speech-alone or speech + gesture instruction to avoid the featural match. In contrast, younger children's visual attention is oriented away from the featural match more effectively by speech + gesture than the speech-alone instruction. This suggests that younger children use the added support of gesture to disambiguate speech and orient their attention away from featural matches.

In sum, the main effect of condition for following score suggests that gesture is effective for directing all children's attention to the referents of spoken instruction. However, when considering the ambiguous portion of instruction, we see differences across age in the relative effectiveness of instruction. For older children, the alignment provided in spoken instruction, "*See, the boy is chasing the girl, and the dog is chasing the cat*" is enough context to recognize that when the instructor refers to the '*boy chasing the girl*' that the boy being referenced is the boy in the chasing relation, not the featural match that is outside of the relation: there is no added benefit of gesture for disambiguating speech. However, for the younger children, we see that gesture *does* have an effect. Children are less likely to look to the featural match when they receive speech and gesture instruction, compared to speech alone instruction. This suggests that gesture is helping disambiguate spoken instruction for these younger children.

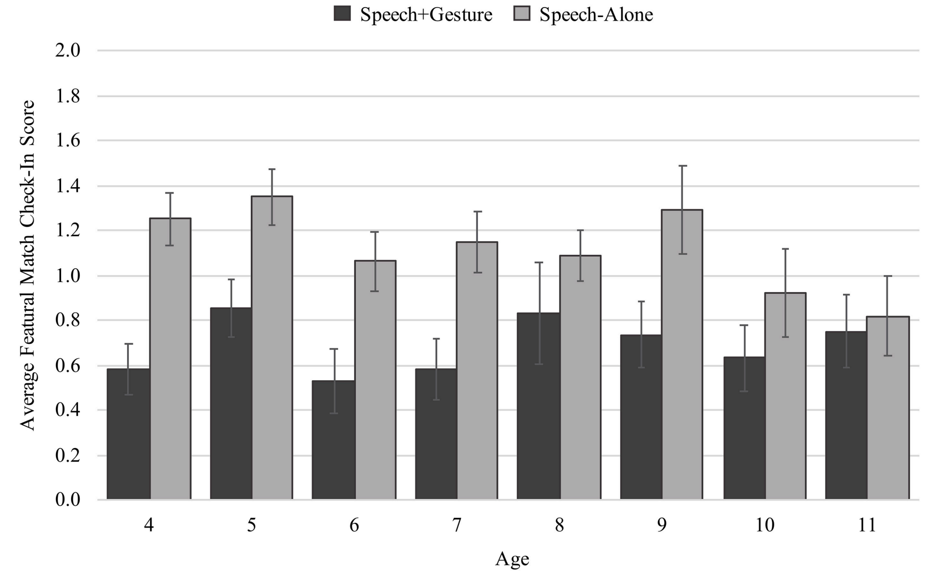


FIGURE 6 | Average check-in scores split by age and condition.

TABLE 1 | *Post hoc* analyses for testing condition effects predicting featural match check-ins.

Age	Beta (*SE*)	*p*-value
4 year-olds	−1.33 (0.32)	<0.001
5 year-olds	−0.92 (0.35)	**0.012**
6 year-olds	−1.07 (0.38)	**0.008**
7 year-olds	−1.13 (0.37)	**0.005**
8 year-olds	−0.51 (0.44)	0.259
9 year-olds	−1.11 (0.47)	**0.024**
10 year-olds	−0.58 (0.46)	0.216
11 year-olds	−0.14 (0.46)	0.770

Bolded p values indicate significant condition effects.

Impact of Visual Attention During Instruction on Children's Analogical Reasoning

Having established that gesture does impact visual attention during instruction, whether this is for all children (following) or only children of particular ages (featural match check-in), we ask whether these patterns of visual attention can explain our behavioral results – that overall speech + gesture seems to marginally improve performance compared to speech-alone, but that this effect is driven by 5-year-old children, who show significantly better performance following speech + gesture instruction compared to speech-alone instruction.

To understand the relation between following along during instruction and performance on the post-instruction trial, we asked whether trial accuracy (0, 1) was predicted by following score. Age was not included in the model, as we found that it was not a relevant predictor of following. Our model revealed that following score was not predictive of accuracy ($\beta = 0.07$, $SE = 0.06$, $t = 1.08$, $p = 0.280$). This suggests that even though gesture helps children follow along with spoken instruction, this organization of visual attention does not contribute to its learning effects in the case of scene analogies.

To understand the relation between checking in with the featural match during instruction and performance on the post-instruction trial, we took into account our finding that, in general, younger children checked in less with the featural match when they received speech + gesture instruction than speech-alone instruction, but older children did not show this difference. This distinctly different pattern of results between younger and older children motivated the use of a median split by age (see Wakefield et al., 2017 for a similar approach): we constructed two models to ask whether check-ins during instruction were predictive of performance on the post-instruction trial for older (8–11 years) and younger (4–7 years) children separately. Here, we found that, whereas older children's check-ins with the featural match did not significantly predict their accuracy at post-instruction ($\beta = -0.12$, $SE = 0.18$, $t = -0.66$, $p = 0.512$), younger children's check-ins with the featural match *were* predictive of their performance on the post-instruction trial: check-ins were negatively related to successful problem solving ($\beta = -0.45$, $SE = 0.18$, $t = -2.58$, $p = 0.009$). This suggests that the ability of gesture instruction to direct attention away from the featural match and disambiguate the meaning of an instructor's speech is the critical factor impacting analogical understanding for younger children.

DISCUSSION

The goals of the present study were to explore whether the impact of adding gesture to spoken instruction on analogical reasoning depends on children's cognitive profile, and to use eye tracking to further understand how gesture might facilitate learning by disambiguating spoken instruction. Our behavioral results suggest a marginal benefit of gesture instruction over speech alone, but only 5-year-old children showed a distinct advantage from speech + gesture instruction when solving the post-instruction trial. This suggests that age – which we demonstrated was a good proxy for cognitive profile based on the relation between performance measures, visual attention, and age, in keeping with previous literature – does impact the utility of gesture for supporting analogical reasoning ability. To understand how disambiguation of speech may play a role in these results, we turned to eye tracking. We found evidence that gesture helps children follow along with spoken instruction, but that this was not predictive of successful problem solving post instruction. Rather, check-in score – visual attention toward the featural match at the point in instruction that was most ambiguous – was negatively predictive of post instruction success for younger children, but not for older children. This lends support to previous arguments that at the root of children's struggle with analogical reasoning is an inability to ignore featural, or superficial, matches in favor of relational matches, and that looking to the featural match is associated with making these types of errors (e.g., Thibaut et al., 2010; Thibaut and French, 2016; Guarino et al., 2019). Although more work must be done to fully explore the impact of gesture instruction for analogical reasoning, these results suggest that one way gesture may help learning in this domain is through directing visual attention in a way that clarifies spoken instruction, but how much of a boost children get depends on their cognitive profile.

Our results suggest that in the case of analogical reasoning, gesture's ability to disambiguate speech may be particularly useful for 5-year-old children who have the foundational cognitive abilities in place to benefit from gesture during instruction. Five-year-old children may be at a pivotal time in development of analogical reasoning ability: while they have a limited cognitive profile and immature analogical reasoning, their inhibitory control and working memory capacity are developed to the point that they can utilize the added support gesture provides. This finding that prior knowledge and ability impacts the utility of gesture corroborates other work in the gesture-for-learning literature. Children need some degree of prior knowledge within a domain that serves as a foundation that gesture instruction can build from (Wakefield and James, 2015; Congdon et al., 2018).

Importantly, our eye tracking data suggest what the added benefit of gesture might be: 5-year-old children showed an increased ability to follow along with instruction and less check-ins with the featural match when they learned through speech and gesture instruction versus speech alone instruction. Thus, the argument could be made that gesture is helping organize children's visual attention in relation to spoken instruction and clarifying ambiguous instruction. But, only check-ins predicted success on the post instruction trial. Considering this in relation to previous work with eye tracking, this may seem puzzling. Wakefield et al. (2018b) found that following along with spoken instruction *did* predict subsequent performance in the case of mathematical equivalence. However, in their measure of following, spoken instruction was ambiguous; whereas in the present study, the general measure of following encompassed spoken instruction that was predominately not inherently difficult for children to decipher because the majority of items referenced in speech could only be associated with one unique item in a scene. In contrast, the speech during the featural match check-in measure *was* ambiguous, and is thus more analogous to the measure of following used by Wakefield et al. (2018b). In both of these cases, gesture is effective at clarifying parts of spoken instruction that are ambiguous, yet critical, for learning. Taken together, results from the current study and previous work suggest that gesture's power to disambiguate spoken instruction is an important mechanism by which gesture shapes learning. And in the case of analogical reasoning, gesture can help children overcome one of the most challenging aspects of problem solving: clarifying for these children which items are in the relation of chasing and critical for solving the analogy, by helping them avoid the lure of a featural match.

While 5-year-olds may be in the developmental 'sweet spot' to benefit from gesture instruction, why does incorporating gesture not benefit all children equally? For all other children, those younger and older than 5 years, there was not a significant benefit of speech and gesture, compared to speech alone instruction, on post-instruction performance. It makes sense that older children (8–11-year-olds) demonstrated learning after both types of instruction: these children seemingly have all the necessary cognitive abilities and prior knowledge needed to utilize either type of instruction. Even though they struggled prior to instruction, their more developed inhibitory control and working memory allowed them to learn even from speech-alone instruction, and the addition of gesture is not necessary for learning the task. This is evidenced by the lack of difference between the number of check-ins with the featural match in the speech + gesture versus speech-alone conditions. Likely because they had the capacity to hold more information in working memory, they were able to consider the instructor's alignment of the chasing relations and recognize which items were being referenced during instruction based on spoken instruction alone, and did not need gesture to organize their visual attention and help them make sense of instruction.

On the other end of the age range, the youngest children, 4-year-old children, may not have a sufficient cognitive profile in place to benefit more from gesture instruction than speech alone instruction. While gesture supports effective visual attention during instruction for these children, their inhibitory control and working memory may be too underdeveloped to extend their understanding beyond the moment, when the support of gesture is no longer immediately present. Thus, even though they looked to the featural match less in the gesture condition, they could not process the multiple relations mentioned in spoken instruction effectively.

Interestingly, 6- and 7-year-old children did not perform similarly to 5-year-old children or older children. While their visual attention was more effectively guided by a combination of speech and gesture instruction, as seen with their younger peers, they did not show the added benefit of gesture post instruction. The non-significant difference between conditions at post-instruction performance for these children may speak to their ability to disambiguate the instructions to some extent when only speech was provided. That is, these children may be able to disambiguate the instructions even with speech alone to a greater extent than 4- or 5-year-olds, but not as effectively as older children. And because they have slightly more mature cognitive profiles (i.e., more developed inhibitory control and working memory) than younger children, they may be better equipped to extend their understanding gained during instruction to post-instruction solving. Together, these results reflect that children's cognitive profile makes a difference for whether gesture facilitates learning above and beyond speech alone instruction.

While this work makes strides toward understanding the nuances of gesture's effects on learning, there are potential limitations that should be addressed. First, we suggest that age can serve as a proxy for a child's cognitive profile without having independent measures of inhibitory control or working memory. Although collecting independent measures of inhibitory control and working memory would have been ideal, previous work using scene analogies has established that inhibitory control and working memory correlate with children's age (5–11-years-old: Simms et al., 2018) and with their analogical reasoning ability over development (working memory: Simms et al., 2018; inhibitory control: Guarino et al., under revision), *and* that children's visual attention is correlated with performance and inhibitory control (Guarino et al., under revision). Specifically, inhibitory control, measured using the Erikson Flanker task, is positively correlated with accuracy and attention to relationally similar items prior to instruction, and negatively correlated with choosing the featural match and attention to the featural match. Therefore, while it may be advantageous in future work to collect direct measures of children's cognitive profile, here, we find the same relation between age, visual attention patterns, and analogical reasoning ability as has been documented in previous work. We are therefore confident that, motivated by previous work, age is associated with cognitive profile.

A second potential limitation is the length of our intervention, which consisted of one pre-instruction trial, two instruction trials, and one post-instruction trial. We designed the study based on previous gesture-for-learning literature showing children *can* benefit from a short intervention (Valenzeno et al., 2003; Church et al., 2004; Rowe et al., 2013). For example, Church et al. (2004) tested children's knowledge of three types of Piagetian conservation (water, length, and number) using one question about each type of conservation before and after they watched one instructional video about conservation that either incorporated speech and representational gestures or speech alone. Similarly in the analogical reasoning literature, Gentner et al. (2016) tested how well children can analogically compare separate contexts after a short intervention. They first exposed children to one pair of model skyscrapers that varied in degree of alignment based on experimental condition, and then asked them build a structure as tall as possible that was 'strong' and repair a structure so it was 'strong.' Through successful comparison of the two model skyscrapers children could identify that a diagonal brace helps make a building 'stronger.' In the present study, we did find an effect of gesture instruction, above-and-beyond that of speech alone instruction, for children at a pivotal point in their analogical reasoning development. This suggests that once again, gesture can impact performance in a short period of time. However, it would be interesting to conduct future work lengthening the period of instruction, as this may allow children more opportunity to benefit from instruction, especially younger children who may need more examples to support their learning.

Finally, while not a limitation, the current work represents a starting, not an ending point, motivating additional questions to answer. For example, similar work using the test-bed of analogical reasoning should consider even younger children. The children in this study likely all had an underdeveloped, but nevertheless present, relevant cognitive profile to support the rudimentary stages of analogical reasoning (e.g., Davidson et al., 2006). Even 4-year-olds have been shown to have some degree of inhibitory control and working memory that allow them to make very simple comparisons – one of the basic building blocks for mature analogical reasoning (e.g., Davidson et al., 2006). To more fully understand the impact of gesture on children with little to no relevant cognitive skills, one could extend and adapt this task to incorporate 2- or 3-year-olds, given that some suggest children younger than 4-years-old have rudimentary relational reasoning capabilities (e.g., Goswami and Brown, 1989; Rattermann and Gentner, 1998; Ferry et al., 2015). The expectation would be that younger children, like 4-year-olds, would not benefit from gesture more than speech alone, and would strengthen the conclusions drawn from the present data.

Additionally, the impact of gesture is not only nuanced in terms of children's current cognitive profile, but many other contextual or situational factors have been cited as playing a role in the effect on learning. For example, the advantage of speech + gesture compared to speech-alone instruction is not always evident in immediate measures at post-instruction, but rather in follow-up measures, from 24 hours (Cook et al., 2013) to 4 weeks (Congdon et al., 2017) after initial training. The one-trial post-instruction assessment may have limited the evaluation of learning.

In sum, the results of the present study extend our understanding of how gesture instruction impacts learning to the domain of analogical reasoning, while providing further insight into how gesture can help disambiguate spoken instruction and how individual differences in a child's cognitive profile impacts the utility of gesture. These findings have important implications for designing teaching methods to support analogical reasoning, but also using gesture as a teaching tool more broadly. Because analogical reasoning shows such a protracted development, due to a slowly developing cognitive profile, it seems that only at certain points will gesture help children more than speech only instruction. Recognizing when this tool can be used could lead to faster growth in a skill that is at the root of a wide range of cognitive skills, such as innovation and creativity (for review see Halford, 1993). More broadly, this work speaks to one of the reasons *why* gesture helps learning, but also emphasizes that individual differences influence the impact gesture can have. Future work should continue to delve into the mechanisms by which gesture shapes learning and consider a child's cognitive state as an important piece of this puzzle.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Open Source Framework at https://doi.org/10.17605/OSF.IO/PAQ4S.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Loyola University Chicago. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

KG and EW: conceptualization, formal analysis, project design and methodology, project administration and supervision, managing resources and software, validation of methodology, writing – original draft, and writing – review and editing. KG: data curation and investigation. EW: funding acquisition. Both authors contributed to the article and approved the submitted version.

FUNDING

Funding for this study was provided by Loyola University Chicago to EW.

ACKNOWLEDGMENTS

We would like to thank Madeline Jurcev, Sasha Stojanovich, Allison Haussler, Natalia Zielinski, Erin Foy, Laura Bottei, Kiril Nikolovski, Dylan Nelson, Victoria Roberts, Tigist Mersha, Gina Wrolstad, and Mariana Sosa for their help with data collection and coding, and the Museum of Science and Industry in Chicago for allowing us to conduct research on site.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.575628/full#supplementary-material

REFERENCES

Alexander, P. A., Willson, V. L., White, C. S., and Fuqua, J. D. (1987). Analogical reasoning in young children. *J. Educ. Psychol.* 79, 401–408. doi: 10.1037/0022-0663.79.4.401

Alibali, M. W., and Nathan, M. J. (2007). "Teachers' gestures as a means of scaffolding students' understanding: evidence from an early algebra lesson," in *Video Research in the Learning Sciences*, eds R. Goldman, R. Pea, B. Barron, and S. J. Derry (Mahwah, NJ: Erlbaum), 348–365.

Altmann, G. T. M., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1

Church, R. B., Ayman-Nolley, S., and Mahootian, S. (2004). The role of gesture in bilingual education: does gesture enhance learning? *Int. J. Bilingual Educ. Bilingualism* 7, 303–319. doi: 10.1080/13670050408667815

Congdon, E. L., Kwon, M. K., and Levine, S. C. (2018). Learning to measure through action and gesture: children's prior knowledge matters. *Cognition* 180, 182–190. doi: 10.1016/j.cognition.2018.07.002

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., Keefe, L. O., and Goldin-Meadow, S. (2017). Better together: simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learn. Instr.* 50, 65–74. doi: 10.1016/j.learninstruc.2017.03.005

Cook, S. W., Duffy, R. G., and Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Dev.* 84, 1863–1871. doi: 10.1111/cdev.12097

Davidson, M. C., Amso, D., Cruess, L., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44, 2037–2078. doi: 10.1016/j.neuropsychologia.2006.02.006

Doumas, L. A. A., Morrison, R. G., and Richland, L. E. (2018). Individual differences in relational learning and analogical reasoning: a computational model of longitudinal change. *Front. Psychol.* 9:1235. doi: 10.3389/fpsyg.2018.01235

Ferry, A. L., Hespos, S. J., and Gentner, D. (2015). Prelinguistic relational concepts: investigating analogical processing in infants. *Child Dev.* 86, 1386–1405. doi: 10.1111/cdev.12381

Gentner, D. (1983). Structure mapping: a theoretical framework for analogy. *Cogn. Sci.* 7, 155–170. doi: 10.1016/S0364-0213(83)80009-3

Gentner, D. (1988). Metaphor as structure mapping: the relational shift. *Child Dev.* 59, 47–59. doi: 10.2307/1130388

Gentner, D. (2010). Bootstrapping the mind: analogical processes and symbol systems. *Cogn. Sci.* 34, 752–775. doi: 10.1111/j.1551-6709.2010.01114.x

Gentner, D., and Smith, L. A. (2013). "Analogical learning and reasoning," in *The Oxford Handbook of Cognitive Psychology*, ed. D. Reisberg (New York, NY: Oxford University Press), 668–681.

Gentner, D., Levine, S. C., Ping, R., Isaia, A., Dhillon, S., Bradley, C., et al. (2016). Rapid learning in a children's museum via analogical comparison. *Cogn. Sci.* 40, 224–240. doi: 10.1111/cogs.12248

Gick, M. L., and Holyoak, K. J. (1980). Analogical problem solving. *Cogn. Psychol.* 12, 306–355. doi: 10.1016/0010-0285(80)90013-4

Glady, Y., French, R. M., and Thibaut, J. P. (2016). "Comparing competing views of analogy making using eye-tracking technology," in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, Philadelphia, PA, 1349–1354.

Glady, Y., French, R. M., and Thibaut, J. P. (2017). Children's failure in analogical reasoning tasks: a problem of focus of attention and information integration? *Front. Psychol.* 8:707. doi: 10.3389/fpsyg.2017.00707

Gordon, P. C., and Moser, S. (2007). Insight into analogies: evidence from eye movements. *Vis. Cogn.* 15, 20–35. doi: 10.1080/13506280600871891

Goswami, U., and Brown, A. L. (1989). Melting chocolate and melting snowmen: analogical reasoning and causal relations. *Cognition* 35, 69–95. doi: 10.1016/0010-0277(90)90037-K

Guarino, K. F., Wakefield, E. M., Morrison, R. G., and Richland, L. E. (2019). "Looking patterns during analogical reasoning: generalizable or task-specific?," in *Proceedings of the Forty-First Annual Meeting of the Cognitive Science Society*, London, 387–392.

Guarino, K. F., Wakefield, E. M., Morrison, R. G., and Richland, L. E. (under revision). Exploring how visual attention, inhibitory control, and co-speech gesture instruction contribute to children's analogical reasoning ability.

Halford, G. S. (1993). *Children's Understanding: The Development of Mental Models*. Hillsdale, NJ: Lawrence Erlbaum.

Huettig, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003

Krawczyk, D. C., Hanten, G., Wilde, E. A., Li, X., Schnelle, K. P., Merkley, T. L., et al. (2010). Deficits in analogical reasoning in adolescents with traumatic brain injury. *Front. Hum. Neurosci.* 4:62. doi: 10.3389/fnhum.2010.00062

Markman, A. B., and Gentner, D. (1993). Splitting the differences: a structural alignment view of similarity. *J. Mem. Lang.* 32, 517–535. doi: 10.1006/jmla.1993.1027

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: The University of Chicago.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., et al. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *J. Cogn. Neurosci.* 12, 260–271. doi: 10.1162/089892904322984553

Morsanyi, K., and Holyoak, K. J. (2010). Analogical reasoning ability in autistic and typically developing children. *Dev. Sci.* 4, 578–587.

Namy, L. L., and Gentner, D. (2002). Making a silk purse out of two sow's ears: young children's use of comparison in category learning. *J. Exp. Child Psychol.* 131, 5–15. doi: 10.1037//0096-3445.131.1.5

Ping, R., and Goldin-Meadow, S. (2008). Hands in the air: using ungrounded iconic gestures to teach children conservation of quantity. *Dev. Psychol.* 44, 1277–1287. doi: 10.1037/0012-1649.44.5.1277

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rattermann, M. J., and Gentner, D. (1998). More evidence for a relational shift in the development of analogy: children's performance on a causal-mapping task. *Cogn. Dev.* 13, 453–478. doi: 10.1016/S0885-2014(98)90003-X

Richland, L. E., Chan, T. K., Morrison, R. G., and Au, T. K. F. (2010). Young children's analogical reasoning across cultures: similarities and differences. *J. Exp. Child Psychol.* 105, 146–153. doi: 10.1016/j.jecp.2009.08.003

Richland, L. E., Morrison, R. G., and Holyoak, K. J. (2006). Children's development of analogical reasoning: insights from scene analogy problems. *J. Exp. Child Psychol.* 94, 249–273. doi: 10.1016/j.jecp.2006.02.002

Rowe, M. L., Silverman, R. D., and Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemp. Educ. Psychol.* 38, 109–117. doi: 10.1016/j.cedpsych.2012.12.001

Simms, N. K., Frausel, R. R., and Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *J. Exp. Child Psychol.* 166, 160–177. doi: 10.1016/j.jecp.2017.08.005

Singer, M. A., and Goldin-Meadow, S. (2005). Children learn when their teachers' gesture and speech differ. *Psychol. Sci.* 16, 85–89. doi: 10.1111/j.0956-7976.2005.00786.x

Starr, A., Vendetti, M. S., and Bunge, S. A. (2018). Eye movements provide insight into individual differences in children's analogical reasoning strategies. *Acta Psychol.* 186, 18–26. doi: 10.1016/j.actpsy.2018.04.002

Thibaut, J. P., French, R., and Vezneva, M. (2010). The development of analogy making in children: cognitive load and executive functions. *J. Exp. Child Psychol.* 106, 1–19. doi: 10.1016/j.jecp.2010.01.001

Thibaut, J. P., and French, R. M. (2016). Analogical reasoning, control and executive functions: a developmental investigation with eye-tracking. *Cogn. Dev.* 38, 10–26. doi: 10.1016/j.cogdev.2015.12.002

Valenzeno, L., Alibali, M. W., and Klatzky, R. (2003). Teachers' gestures facilitate students' learning: a lesson in symmetry. *Contemp. Educ. Psychol.* 28, 187–204. doi: 10.1016/S0361-476X(02)00007-3

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., and Knowlton, B. J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychol. Aging* 19, 581–591. doi: 10.1037/0882-7974.19.4.581

Wakefield, E., Novack, M. A., and Goldin-Meadow, S. (2017). ?Unpacking the ontogeny of gesture understanding: how movement becomes meaningful across development. *Child Dev.* 89, 245–260. doi: 10.1111/cdev.12817

Wakefield, E. M., Hall, C., James, K. H., and Goldin-Meadow, S. (2018a). Gesture for generalization: gesture facilitates flexible learning of words for actions on objects. *Dev. Sci.* 21:e12656. doi: 10.1111/desc.12656

Wakefield, E. M., and James, K. H. (2015). Effects of learning with gesture on children's understanding of a new language concept. *Dev. Psychol.* 51, 1105–1114. doi: 10.1037/a0039471

Wakefield, E. M., Novack, M. A., Congdon, E. L., Franconeri, S., and Goldin-Meadow, S. (2018b). Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci.* 21:e12664. doi: 10.1111/desc.12664

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 23 September 2020
doi: 10.3389/fpsyg.2020.01935

What's New? Gestures Accompany Inferable Rather Than Brand-New Referents in Discourse

Sandra Debreslioska[1]* *and Marianne Gullberg*[1,2]

[1]Centre for Languages and Literature, Lund University, Lund, Sweden, [2]Lund University Humanities Lab, Lund University, Lund, Sweden

The literature on bimodal discourse reference has shown that gestures are sensitive to referents' information status in discourse. Gestures occur more often with new referents/first mentions than with given referents/subsequent mentions. However, because not all new entities at first mention occur with gestures, the current study examines whether gestures are sensitive to a difference in information status between brand-new and inferable entities and variation in nominal definiteness. Unexpectedly, the results show that gestures are more frequent with inferable referents (hearer new but discourse old) than with brand-new referents (hearer new and discourse new). The findings reveal new aspects of the relationship between gestures and speech in discourse, specifically suggesting a complementary (disambiguating) function for gestures in the context of first mentioned discourse entities. The results thus highlight the multi-functionality of gestures in relation to speech.

Keywords: gestures, discourse, reference, information status, speech-gesture relationship

OPEN ACCESS

Edited by:
Kazuki Sekine,
Waseda University, Japan

Reviewed by:
Wing Chee So,
The Chinese University of Hong Kong, China
Elena T. Levy,
University of Connecticut, United States

Correspondence:
Sandra Debreslioska
sandra.debreslioska@ling.lu.se

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

Received: *28 April 2020*
Accepted: *13 July 2020*
Published: *23 September 2020*

Citation:
Debreslioska S and Gullberg M (2020) What's New? Gestures Accompany Inferable Rather Than Brand-New Referents in Discourse. Front. Psychol. 11:1935. doi: 10.3389/fpsyg.2020.01935

INTRODUCTION

When producing a stretch of discourse, speakers can use speech and speech-associated gestures to indicate to whom or what they are referring. Bimodal referring is a widely acknowledged phenomenon, but the mechanism explaining why gestures occur at specific moments when speakers mention entities in discourse is less well understood. McNeill (1992, 2005) proposes that communicative dynamism (CD) – the degree to which a piece of information "pushes the communication forward" (Firbas, 1971, p. 136) – determines the presence versus absence of gesture. McNeill takes information status, one of three factors influencing CD (Firbas, 1971), as a starting point and shows that the less accessible the information, the more likely a gesture is to occur. Conversely, the more accessible the information, the less likely a gesture is to occur. This would suggest that new entities in discourse are more likely to occur with gestures than already mentioned ones, an observation that is well supported in the literature (Marslen-Wilson et al., 1982; Levy and McNeill, 1992; McNeill and Levy, 1993; Gullberg, 1998, 2003, 2006; Levy and Fowler, 2000; Foraker, 2011).

However, there is evidence that not all entities which are mentioned for the first time in discourse, representing the lowest degree of accessibility (or highest degree of newness), are accompanied by gestures (e.g., Gullberg, 2003; Foraker, 2011). Hitherto, this variation has gone unmentioned. The current study therefore examines the variation in the incidence of gesture with entities mentioned for the first time and specifically probes the possibility that gesture production may be related to entities' information status (brand-new vs. inferable; Prince, 1981; see also Clark, 1977; Fillmore, 1982; Chafe, 1994; Givón, 1995; Gundel, 1996), which in turn may interact with nominal definiteness [definite vs. indefinite noun phrases (NPs)].

Speech-Associated Gestures

When speakers engage in talk, bodily action is always mobilized, which goes beyond the use of the anatomical apparatus needed for speaking (Kendon, 2014). This bodily action can involve the face and eyes, the neck and head, the upper body and trunk, and importantly, the hands and arms. A large body of research shows that the hand and arm movements speakers perform while speaking (also called gesticulations, co-speech gestures, speech-associated gestures, manual gestures, or simply gestures) are organized as patterns of movement that are rhythmically coordinated with speech production (Kendon, 1972, 1980). At the same time, they are also considered to be meaningful, specifically in how they relate to the meanings in the speech they accompany (McNeill, 1992; Kendon, 2004). For instance, speakers may use gestures to provide iconic representations of what is being talked about or they may use them to point to or locate entities (see **Figures 1, 2**). In **Figure 1**, the speaker mentions the entity *Ärmel* "sleeve" for the first time within a sewing event. In exact temporal co-occurrence with this mention, she uses a gesture to represent the sewing action performed on the sleeve by moving her right hand in a circular fashion along her left arm producing an iconic depiction. In **Figure 2**, the speaker mentions the existence of the entity *Tisch* "table" for the first time. She raises both hands in parallel from her lap to about chest level, with flat hands and palms facing each other, in order to indicate the shape/size of the table. This tight coordination in meaning and timing of two modalities is at the basis of the consideration that gestures and speech are conceptually linked (Kendon, 2004).

Speech-Associated Gestures and the Information Status of Entities

The relationship between speech and gestures extends from the local level of one composite expression to more global interactions of the two modalities, as is the case for the organization of connected discourse. Gestures and speech vary in a coordinated fashion in the way they are deployed depending on the unfolding of information in discourse. For example, for the tracking of referents in discourse, a growing number of studies demonstrate a close link between gestures and speech, emphasizing the role played by the information status of entities. When entities are new or less accessible, they are typically expressed with richer referential expressions in speech (as in lexical NPs) and are accompanied by gestures. In contrast, when entities are given or more accessible, they are expressed with leaner referential expressions in speech (as in pronouns) and are typically not accompanied by gestures (e.g., Marslen-Wilson et al., 1982; Levy and McNeill, 1992; McNeill, 1992, 2005; McNeill and Levy, 1993; Gullberg, 1998, 2003, 2006; Levy and Fowler, 2000; Yoshioka, 2008; Wilkin and Holler, 2011; Parrill, 2012; Debreslioska et al., 2013; Perniss and Özyürek, 2015; Debreslioska and Gullberg, 2019; but see So et al., 2010 for different results when using

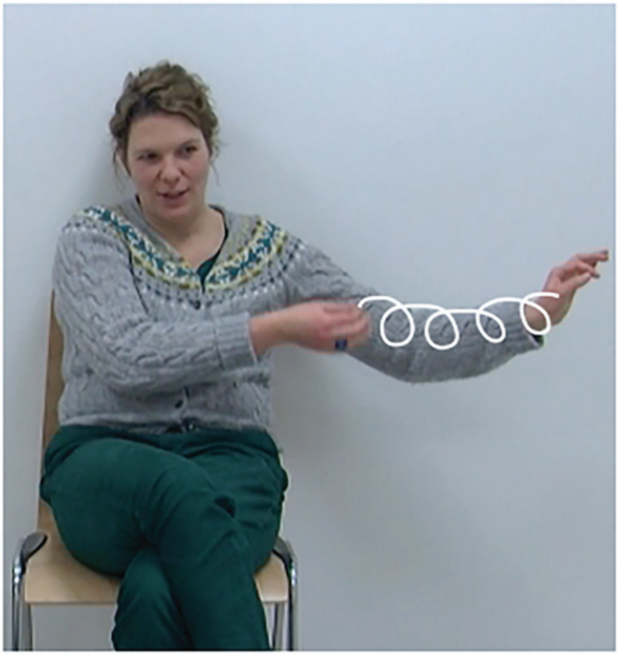


FIGURE 1 | Iconic representation of "sewing a sleeve" (gesture alignment indicated in bold face).
*Wie sie zuerst auf der Seite, auf der die Fee steht, **den Ärmel zun**äht*
"How she **sews the sleeve** on the side, on which the fairy is standing"

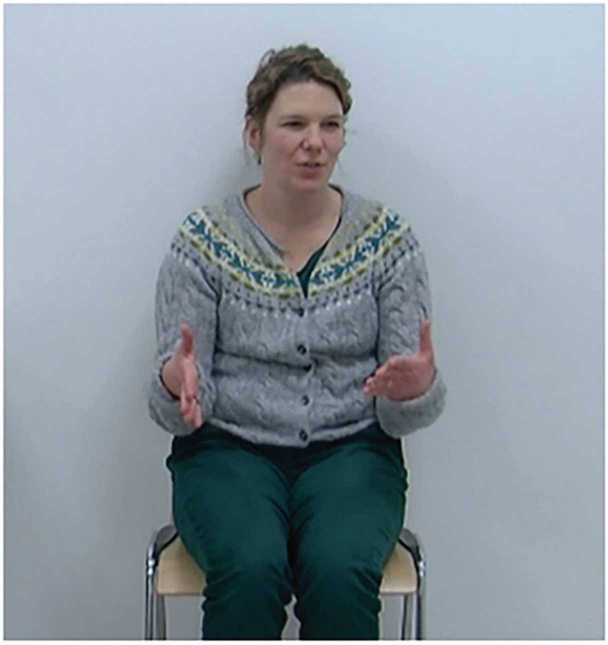


FIGURE 2 | Iconic representation of "a table" (gesture alignment indicated in bold face).
*Und es gibt **ein Tisch***
"And there is **a table**"

a different gesture coding approach). This pattern reflects Givón's so-called principle of quantity (Givón, 1983), which predicts more marking material for less accessible information and less marking material for more accessible information (see also Ariel, 1988, 1991, 1996; Prince, 1992; Gundel et al., 1993; Chafe, 1994; Arnold, 1998, 2008, 2010; and for child discourse, see e.g., Clancy, 1993; Hickmann and Hendriks, 1999; Allen and Schroder, 2003; Narasimhan et al., 2005; Serratrice, 2005; Allen, 2008). More importantly, the pattern is also at the heart of McNeill's theory of CD and gestures, which posits that the more a piece of information "pushes the communication forward" (Firbas, 1971, p. 136), the more likely it is that a gesture co-occurs with it. The information status (or how accessible a referent is) is one important factor influencing the CD of an expression (Firbas, 1971). Findings showing the parallelism between speech and gesture to signal new information (richer referential expressions and gestures) versus given information (leaner referential expressions and few/no gestures) are considered to be support for McNeill's theory.

An example of this pattern is illustrated in (1), taken from the data set of the current study. In order to signal that referents are new, indefinite lexical NPs are used in speech for the referents *Kerzen* "candles" in utterance 1, and *Fee* "fairy" in utterance 2. When the referent "candles" is mentioned for the second time in utterance 2, the speaker uses a pronoun to refer back to it (*die* "they"). In gesture, this alternation between richer/leaner expressions is reflected in a variation in gesture incidence. Both first mentions are accompanied by gestures (i.e., the referents "candles" and "fairy," marked in bold face), but the subsequent mention of the referent "candles" by the pronoun *die* "they" is not.

(1)

1 *Und auf der Torte ähm sind **Kerzen$_1$** drauf* "and on the big cake are candles."
2 *Die$_1$ werden angezündet von ähm **einer Fee$_2$*** "they are being lit up by a fairy."

Although the literature thus shows that new referents are more likely to occur with gestures than old/given ones, it also shows that not all first mentions are accompanied by gesture (e.g., 39.8% in Foraker, 2011; 75% in Gullberg, 2003). This observation, in turn, seems to challenge predictions derived from McNeill (1992, 2005). Since a referent mentioned for the first time should always push the communication forward (or carry higher CD), we might expect every first mention to be accompanied by gesture. But it is not. It remains unclear why this should be the case.

One possibility is that a more fine-grained notion of information status is needed to account for the incidence of gestures. Specifically, in the context of new information and first mentions, entities could be divided into those that are brand-new and those that are inferable from the preceding context. Prince (1981, 1992) defines brand-new entities as being new to the preceding discourse and also new to the addressee. Inferable entities, on the other hand, are new to the preceding discourse, but their existence can be inferred by the addressee. A referent is typically rendered inferable by virtue of a trigger entity, which has previously been mentioned in the discourse (Prince, 1981, 1992). For instance, inferable referents are entities that stand in a part/whole relationship or in a content/container relationship to already-mentioned entities. For example, if the referent *Besen* "broom" has already been mentioned in a particular stretch of discourse, then a current mention of the referent *Stiel* "broomstick" can be considered inferable. Similarly, if the referent *Salzstreuer* "saltshaker" has already been mentioned, then a current mention of *Salz* "salt" can be considered inferable information. Note that these kinds of relationships that give rise to inferables hold true even if in some cases a particular referent does not have a certain part or content (e.g., an empty saltshaker). It is considered sufficient that the relationship typically holds true (Birner, 2013). More recent accounts further argue that inferable information should rather be regarded as "hearer new" but "discourse old" (Birner and Ward, 1998; Birner, 2013). This view emerges from observations that inferable information is often used in sentence constructions which depend on "discourse old" information on the one hand and in constructions which depend on "hearer new" information on the other.

The variation in information status between brand-new versus inferable referents can be signaled in speech by a formal variation in nominal definiteness. Speakers are likely to refer to inferable entities with definite lexical NPs (also called bridging expressions, as in e.g., *the broomstick*) more often than to brand-new entities (e.g., *a broom*; Clark, 1975, 1977). In principle, however, inferable entities can be represented by both indefinite and definite lexical NPs (Prince, 1992; Gundel, 1996), as illustrated in examples (2–3), taken from the current data set. In each case, the speaker has already introduced a broom as a whole into the discourse. At a later point, one speaker mentions the referent "broomstick" by using an indefinite lexical NP (2), whereas the other speaker chooses a definite lexical NP (3). In order to avoid circularity (i.e., by assuming that each definite nominal used for a first mention automatically represents an inferable entity, and vice versa), we will keep the formal marking of nominal definiteness separate from information status while still assuming that the two measures will co-vary, such that inferables will be referred to with definites more often.

(2) PP1: *Der hat <u>nen braunen Stiel</u> und gelbe Borsten* "it has <u>a brown broomstick</u> and yellow bristles."
(3) PP8: *<u>Der Besenstiel</u> ist braun* "<u>the broomstick</u> is brown."

McNeill's (1992, 2005) theory of CD and gesture, but also most other previous research on gestures in discourse, would predict that brand-new referents – which are "truly" new since they have never been mentioned and cannot be inferred from previously mentioned referents – should attract more gestures than inferable referents. Furthermore, if it is the case that indefinite lexical NPs signal brand-new referents more than definite lexical NPs, then they should also attract gestures more than definite lexical NPs (Debreslioska and Gullberg, 2019; but see Wilkin and Holler, 2011).

The current study seeks to test these predictions in order to further our understanding of when first mentions attract gestures or not.

The Current Study

The current study examines when discourse entities that are mentioned *for the first time* co-occur with gestures and when they do not. Particularly, it explores two variables, information status (brand-new vs. inferable reference) and nominal definiteness (definite vs. indefinite nominals) to test whether these two factors are related to the incidence of gestures (presence vs. absence) in bimodal discourse.

For speech, we predict that (a) brand-new entities are more likely to be mentioned with indefinite nominals, and conversely, that inferable entities are more likely to be represented with definite nominals. For gesture, we predict that, if information status and definiteness have an effect on the incidence of gestures, (b) brand-new referents will co-occur with gestures more than inferable referents, and (c) indefinite lexical NPs will co-occur with gestures more than definite lexical NPs.

MATERIALS AND METHODS

Participants

We invited 20 native German speakers (16 female, mean age = 26, range 20–39) to participate in the study at Ludwig-Maximilian University, Munich, Germany. All participants came with a native German-speaking friend who acted as listener. Everyone provided written consent.

Materials/Design

We used a picture story to elicit narrative speech and gestures. The story consisted of 127 pictures about three fairies, each having to fulfill a task (baking a cake, sewing a dress, and cleaning the floor), which they fail at, and consequently use magic to achieve (see **Figures 3–5** for examples). References to the three fairies and a range of inanimate entities were considered (see **Appendix B** for a full list).

Procedure

Participants sat across from each other and only the speaker was captured by a video camera, focusing on head and torso. Participants read instructions on paper, and the experimenter further repeated the main points orally to them. Speakers had to retell the picture story by answering the question "what happened?" Since the story was rather long, speakers only saw four to six pictures at a time, had unlimited time to memorize them, and then retold that piece to the listener before moving on to the next one. Speakers were encouraged to say something about each picture. The listener was not supposed to ask any questions, but to write down a short summary of each part of the story they just heard. While only the speaker was of interest for the current study, this was not disclosed to the participants. The listener was also instructed not to cross legs or arms in order to avoid mirroring by the speaker, which could be unfavorable for gesture production (e.g., Kendon, 1973; Chartrand and Bargh, 1999). The roles of speaker and listener were assigned randomly[1]. A session lasted between 45 and 90 min. The produced narratives were 20 min long on average.

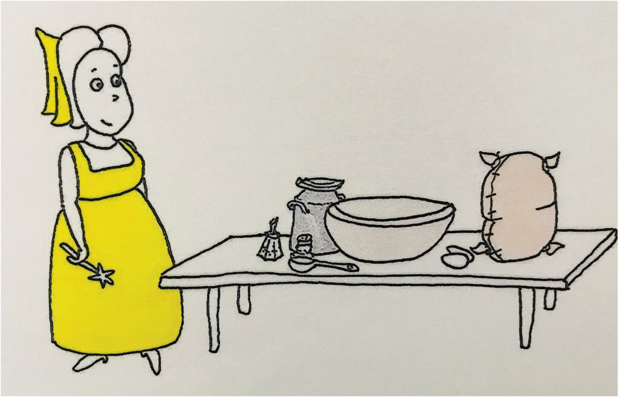


FIGURE 3 | Example stimulus picture 1.

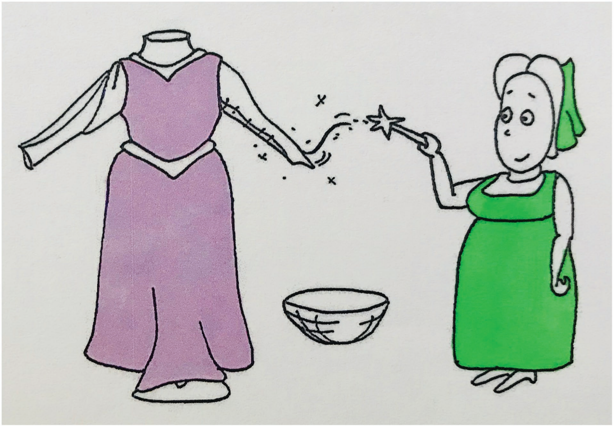


FIGURE 4 | Example stimulus picture 2.

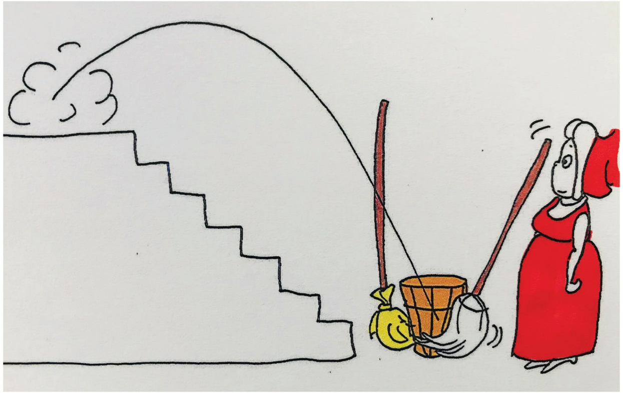


FIGURE 5 | Example stimulus picture 3.

[1] However, if one of the participants knew that the experimenter researched gestures (e.g., if a research assistant from the local university working on the topic of gestures came with a friend), then she was automatically assigned as listener.

All participants were debriefed orally at the end of the experiment and were offered refreshments as compensation. Furthermore, all participants signed consent forms; while speakers also filled out a more detailed (language) background questionnaire based on work of Gullberg and Indefrey (2003).

Speech Coding

A native speaker of German transcribed speech of all 20 narratives produced by the participants using German standard orthography, also taking note of filled pauses, word truncation, repetitions, etc. We then identified all referential expressions mentioning an entity for the first time. For the purposes of this study, we only selected references to concrete animate (i.e., the fairies) and inanimate entities (e.g., cake, broom, needle; see **Appendix B** for a full list of entities) that played a role in the story. We excluded all references to abstract/non-spatial/immaterial objects (as in 4). We also excluded references to "non-referential referents" (Chafe, 1994). Non-referential referents do not factually exist at the moment of mention, and speakers typically mention them in an irrealis context or present their existence as "hypothesized, predicted, or denied" (Chafe, 1994; example 5). Importantly, non-referential referents are not trackable and, thus, represent a different category of referents than those that are to be explored in the present study. Finally, we also excluded references to the pictures themselves (as in 6).

(4) *Sie hat eine Idee* "she has an idea."
(5) *das soll vielleicht so ein Mehlsack sein* "it should perhaps be a bag of flour."
(6) *Die grüne äh steht in der Mitte des Bildes* "the green fairy stands in the middle of the picture."

Entities were either mentioned as core arguments (subjects and direct objects) in presentative utterances (such as existentials or locatives), transitive or intransitive clauses (corresponding to 60% of all referential expressions; see **Table 1**). In all three of these utterance types, the starting point is typically either an inanimate or animate locational element, the dummy subject *es* "it," or the adverbial *da* "there," and the first mentioned entities are placed toward the end of the utterance. In intransitive utterances, the speakers further use subject-verb inversion in order to place the first mentioned entity toward the end of the utterance. Placing the referents in utterance final (focal) position is typical in the context of first mentions. The rest of the entities were instantiated as either oblique arguments (29% of all referential expressions) or in verbless utterances (11% of all referential expressions; **Table 1**; for a construction type analysis and how different constructions are related to representational gestures, see Wu, 2018; Debreslioska and Gullberg, 2020).

Information Status

For each referential expression, we determined whether it referred to a brand-new or inferable entity. A brand-new entity was a "truly" new entity, which had never been mentioned before, and was not inferentially linked to a previous entity in the discourse. Conversely, an inferable entity corresponded to an entity that was mentioned for the first time, but that was linked to a previous "trigger" entity in the discourse *via* an inferential link (following Prince, 1981, 1992). In the current data set, two different links connected first mentions to previous entities, namely part/whole (e.g., sleeve – dress, egg shells – eggs), and content/container relationships (e.g., milk – milk can, sugar – sugar bowl; see **Appendix B** for a full list).

In relation to the way that entities were embedded in different utterance types, we observed that brand-new entities were introduced as core arguments in 67% of the cases, as oblique objects in 21% of the cases, and in verbless utterances in 12% of the cases. Inferable entities were mentioned as core arguments in 41% of the cases, as oblique objects in 50% of the cases, and in verbless utterances in 9% of the cases.

Noun Phrase Definiteness

We considered lexical NPs to be indefinite if they were mentioned as bare nouns, marked by indefinite determiners or numerals (*Milch/ein Besen* "milk/a broom"; *drei Feen* "three fairies"). We considered them to be definite when they were marked by definite determiners, such as definite articles, demonstrative pronouns and possessive pronouns (*die/diese Fee* "the/that fairy"; *ihr Kleid* "her dress").

Gesture Coding

We used frame-by-frame analysis of digital video in the software ELAN (Sloetjes and Wittenburg, 2008) to annotate manual gestures. We identified the most meaningful part of the gestural movements, the stroke phase (McNeill, 1992; Kendon, 2004), with sound turned off. We turn the sound off during the annotation of gesture phases to provide an objective and replicable annotation based on physical features of the hand/arm movements alone. We determined the onset and offset of a stroke when there were

TABLE 1 | Clause types used to introduce referents and examples.

Clause types	Examples
Presentative clauses (existentials; locatives)	*und in dieser Schüssel sind drei Zauberstäbe* "and in the bowl are three wands"
	es gibt einen Tisch "there is a table"
	da sind drei Feen "there are three fairies"
	die hat n Eimer "she has a bucket"
Transitive clauses	*sie holt ein kleines Kästchen* "she goes to get a little box"
Intransitive clauses	*da kommen Funken raus* "there are coming out sparks"
	dann fliegt ein Streichholz herbei "then flies by a match"
Oblique arguments	*in einer Schüssel, hat sie die Kerzen* "in a bowl, she has the candles"
	die eine läuft zum Tisch "one of them walks to the table"
Verbless utterances	*und zwar mit roten Herzchen* "and namely/that is with red hearts"[1]
	und dann das Unterteil "and then the lower part"[2]

[1] *Context of this verbless utterance: PP13: Also die Tube mit dem Zuckerguss ähm verziert den Kuchen dann noch weiter und zwar mit roten Herzchen. "So the icing bag continues to decorate the cake. And namely/that is with red hearts."*
[2] *Context of verbless utterance: PP22: Aber die Nadel näht noch einmal das Oberteil besser zusammen und dann das Unterteil. "But the needle sews together the upper part more appropriately. And then the lower part."*

changes in the trajectory or movement of the hand(s), as well as when there were changes in the tension, shape, or placement of the hand(s) (see Kendon, 2004; Seyfeddinipur, 2006 for more detailed descriptions/instructions). In the case of deictic gestures, we counted the accelerated movement toward the end configuration together with the momentary stop in the end configuration as the stroke. For all other gestures, we also included post stroke holds, defined as movement cessations of the hand at the end of a gesture stroke, as meaningful parts of the gesture. One of the functions of post stroke holds is to allow for the rest of the co-expressive speech to be uttered before the hand goes into retraction or the next gesture (Kita, 1990; McNeill, 1992). They are therefore relevant for our analysis. Since the goal of the current examination is to find out when gestures are aligned with new referents in discourse, it is crucial to take into consideration the full chunk of speech that the meaningful part of the gesture is related to. In a last step, we identified which gestures co-occurred temporally with at least one syllable of the relevant referential expressions (following Stam, 2006; Gullberg et al., 2008) and only took those gestures into account for the analyses.

Reliability Coding

A second German native speaker recoded speech for information status (brand-new vs. inferable) and nominal definiteness (indefinite vs. definite) for the referential expressions of four participants, corresponding to about 20% of the total amount of referential expressions used in the analyses. The agreement between coders was 90% for the coding of information status (brand-new vs. inferable). Interrater reliability was computed using Cohen's kappa (Kappa =0.796, *SE* of kappa =0.035). The agreement between coders was 98% for nominal definiteness coding (indefinite vs. definite nominals). The interrater reliability was also measured using Cohen's kappa (Kappa = 0.979, *SE* of kappa = 0.012).

A second coder recoded gestures for the same four participants in our data set, identifying gestures in the target utterances (i.e., those containing first mentions). The target gestures in those utterances constitute about 20% of the total amount of gestures that went into the analysis. Agreement was reached when the gesture that coder 2 identified aligned with the same referential expression as the one that coder 1 identified. The agreement between coders was 96%.

Analyses

The analyses focus on first mentions of referents, brand-new or inferable, encoded by definite or indefinite nominals and produced with or without gestures. The data set consisted of 1,489 spoken referential expressions and 811 gestures produced by all 20 participants.

We used linear mixed effects models with the lmerTest package (Kuznetsova et al., 2017) in RStudio (RStudio Team, 2016) for all analyses. **Table 2** summarizes the two main analyses. Analysis 1 concerns speech alone, examining the relationship between the information status of referents and their formal representation in speech as definite versus indefinite nominals. Analysis 2 then examines whether the presence of gesture is modulated by these variations in information status and definiteness.

RESULTS

Speech

In a first step, we explored the relationship between information status and definiteness in speech alone (**Table 2**, analysis 1). **Figure 6** presents the observed distribution of indefinite nominals across brand-new (82%) and inferable referents (27%). The analyses revealed that, as expected, brand-new referents were significantly more likely to be expressed as indefinite nominal expressions than inferable referents (*EST* = −5.83, *SE* = 0.32, *z*-value = −18.46, *p* = 0.000). Conversely, inferable referents were significantly more likely to be encoded with definite than with indefinite nominal expressions (*EST* = 4.33, *SE* = 0.30, *z*-value = 14.43, *p* = 0.000).

Gesture

Next, we examined the relationship between the incidence of gestures and first mentions. We found that speakers produced gestures for 55% (*SD* = 23%) of all first mentioned entities (mirroring 60% in Foraker, 2011). We tested whether the incidence of gesture is modulated by two independent variables, namely information status operationalized as brand-new versus inferable, and referents' representation in speech as indefinite versus definite nominals (**Table 2**, analysis 2). **Figure 7** presents the observed distribution of (mean proportions of) gestures across inferable (65%) versus brand-new referents (52%). **Figure 8** presents the observed distribution of (mean proportions of) gestures across definite (56%) versus indefinite (54%) nominals.

We ran five different models in order to determine the model that fit the data best. The first model included no

TABLE 2 | Variables and levels.

Analysis	Dependent variable	Levels	Predictor variable	Levels
1	Definiteness	Indefinite/ Definite	Information status	Brand-New/ Inferable
2	Presence of gesture	yes/no	Information status	Brand-New/ Inferable
			Definiteness	Indefinite/ Definite

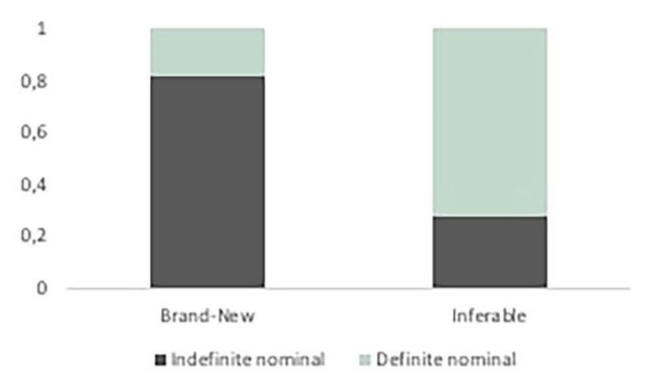


FIGURE 6 | Indefinite nominals representing brand-new versus inferable entities (observed data).

predictor variables. The second and third models each included only one predictor variable, information status and definiteness, respectively. Finally, the fourth and fifth models included both predictor variables, but one was a simple model and the other an interaction model. All models included "subject" as a random predictor variable. We compared the Akaike information criterion (AIC) values between all models in order to determine the model which represented the best fit to the data set. Lower AIC values correspond to better fit (see **Appendix A** for a full list of models ranked according to their AIC values). More specifically, the AIC is an estimate of predictive accuracy, which measures how well a regression model will fit when applied to a new sample (see Long, 2012 for a detailed description).

The model comparisons showed that the simple model including the two predictor variables, information status and definiteness, best explained the present data. The analysis revealed that there was a significant effect of information status on the incidence of gestures but in the opposite direction from the prediction. Inferable referents were significantly more likely to occur with gestures than brand-new referents (*EST* = −0.73, *SE* = 0.16, *z*-value = −4.51, *p* < 0.000). There was no significant effect of definiteness (*EST* = −0.25, *SE* = 0.15, *z*-value = −1.68, *p* = 0.092).

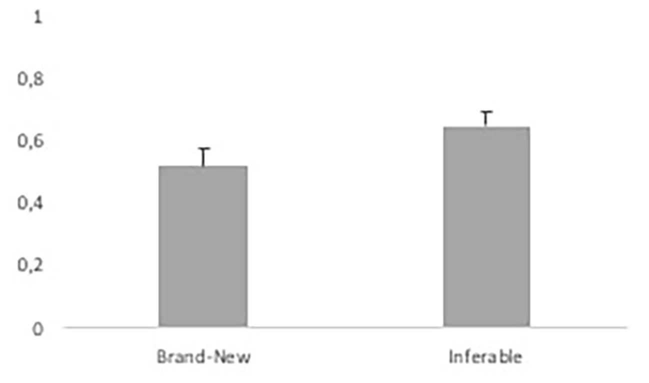


FIGURE 7 | Mean proportions of gestures used with brand-new referents (0.52; *SE* = 0.05) versus inferable referents (0.65; *SE* = 0.5; observed data).

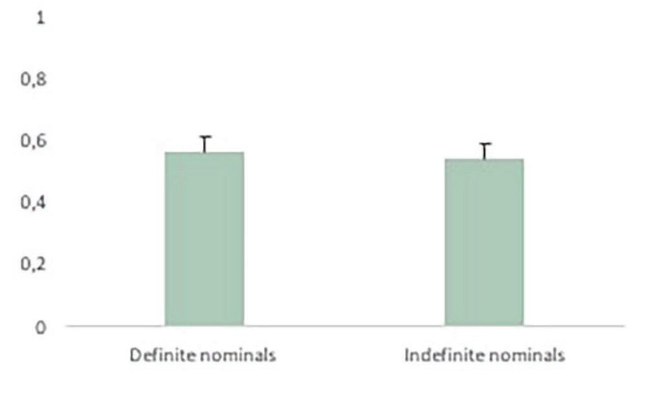


FIGURE 8 | Mean proportions of gestures used with indefinite (0.54; *SE* = 0.5) versus definite nominals (0.56; *SE* = 0.05; observed data).

DISCUSSION

The existing literature on discourse reference and gestures has shown that gestures are sensitive to referents' information status in discourse such that they occur more often with new referents/first mentions than with given referents/subsequent mentions. However, because not all new entities are gestured about at their introduction, the current study set out to examine when first mentions of discourse entities are accompanied by gestures and when they are not. In particular, we considered the possible connection between gesture production and a more fine-grained difference in information status between brand-new and inferable entities, as well as the variation in linguistic encoding between indefinite and definite nominals, reflecting this difference in speech.

The results can be summarized in two points. First, the speech results showed that, as predicted, brand-new referents tend to be expressed by indefinite nominals (e.g., *a broom*), whereas inferable referents tend to be expressed by definite nominals (e.g., *the broomstick*). These findings are in line with previous research on this topic (e.g., Clark, 1975, 1977; Prince, 1981, 1992; Fraurud, 1990; see also Hickmann et al., 1996, for marking of newness in German), showing that referential form is sensitive to the inferability of referents mentioned for the first time.

Second, the gesture results revealed a link between gesture production and the brand-new/inferable distinction. Contrary to prediction, however, inferable referents were significantly more likely to be accompanied by gestures than brand-new ones. For example, the brand-new referent "dust pile" is introduced in a presentative utterance, *man sieht da vorne dran sonen kleinen Haufen* "one sees there in front a little pile," and no gesture co-occurs with this first mention. Compare this to the first mention of the inferable referent "egg yolk" in the presentative utterance *und man sieht jetzt das Eigelb* "and one sees now the egg yolk," in which a gesture localizing the egg yolk above a bowl accompanies the referential expression denoting it (**Figure 9**). In this example, the speaker raises her hand from her lap to about chest level while also using a marked hand shape to represent the shape of the egg yolk. **Figure 9** illustrates the end position of her gesture.

This result poses a challenge to McNeill's (1992, 2005) theory of CD and gestures, which posits that the more a piece of information pushes the communication forward, the more likely it is to co-occur with a gesture. It seemed plausible to assume that brand-new referents, which mark the lowest degree of accessibility of referents in discourse, push communication forward more than inferable referents and would thus be accompanied by gestures more often. However, the current results do not support this assumption.

The study asked whether information status plays a role for the incidence of gestures with first mentioned entities in discourse. The current results suggest that this is the case: gestures are significantly more likely to occur with inferable than with brand-new referents. Although these results go in unanticipated directions, they still suggest that gesture production is sensitive to the subtle distinction in information status suggested by the difference between brand-new and inferable referents. The findings therefore generally support previous research on the relationship between information status and gesture production in discourse (e.g., Marslen-Wilson et al., 1982; Levy and McNeill, 1992; McNeill and Levy, 1993; Levy and Fowler, 2000; Gullberg, 2003, 2006; Foraker, 2011). However, the question is why gestures should be more strongly linked to inferable than to brand-new entities. Birner (2013) proposes that inferable information is "discourse old" but "hearer new." That is, inferable information can be considered "discourse old" because it is inferentially linked to the previous discourse in some way. But it is also "hearer new" because the information itself might not yet be active in the addressee's representation of the discourse (even if in principle it might be more easily accessed than brand-new information). We suggest that speakers may use gestures to highlight these (inferable) pieces of information in order to signal to the addressee that, even if the information is marked by a definite determiner, they are still to add it as a new referent to the discourse representation. In other words, since inferable entities are linguistically encoded similarly to given information (by definite nominals), speakers may produce gestures more often with them to signal to the addressee that the information is not in fact given, but new since there is not yet any active representation of the information in the discourse model. By this account, gestures and speech in this particular context seem to work together in a complementary rather than a parallel fashion. That is, when speech does not provide an unambiguous cue as to whether information needs to be newly added to the discourse representation (such as by indefinite nominals), gestures can do so instead.

This interpretation is something of a departure from previous studies, which have mainly emphasized that the two modalities work in parallel. However, the interpretation is commensurate with McNeill's (1992) view on gestures and speech as two dimensions of the same idea unit, where gestures do not always represent the same information as speech. The suggestion is that together, speech and gesture form a more complete representation. Similarly, Kendon (2014) suggests that gestures and speech together form a richer and more complex expression than if words or gestures are considered alone. In order to form such a complex expression, gestures can be used in flexible ways, as complements or supplements, sometimes even as substitutes or alternatives, to spoken expressions, always in accordance with the underlying communicative effort or intent (Kendon, 1986). The two modalities can thus be seen as adaptable resources allowing speakers to vary how they coordinate them depending on the communicative needs in different types of situations (Gullberg, 1998; Holler and Beattie, 2003; Kendon, 2004).

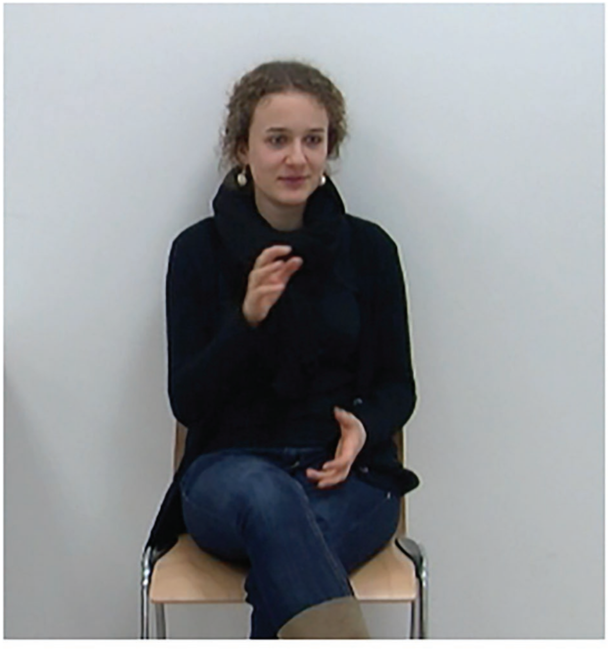


FIGURE 9 | Example of a gesture accompanying the first mention of an inferable entity.
*und man sieht jetzt das Ei**gelb***
"and one sees now the egg **yolk**"

Interestingly, the results can also be related to qualitative observations in children's speech and gesture production. Allen (2008) examined the influence of a referent's information status on children's argument realization in Inuktitut, a pro-drop language. She found that while children predominantly realize an argument overtly when it is "new," there are still surprisingly many cases when children simply drop the argument even if the referent is new to the discourse. Qualitative analyses of some of the cases revealed that those elided arguments often refer to inferable referents instead of brand-new ones suggesting that children seem to differentiate between the two. More interestingly, Allen further showed that children tend to produce a gesture in place of the elided argument (while the timing of the gestures is unclear, we assume that gestures aligned with the verb phrases of an utterance; see also Yoshioka, 2005). That is, when referents represent new, but inferable information, children can drop the argument in speech and use a manual gesture instead. Often, this would be a deictic gesture pointing to the intended referent. Therefore, Allen's (2008) analyses similarly suggest that when new but inferable information is linguistically treated like given information (i.e., by zero arguments in Allen's study; by a definite determiner in the present study), a gesture might indicate the referent's accessibility instead.

Importantly, although referent inferability explains a considerable part of the data, we still find inferable referents that are not accompanied by gestures (36%), as well as brand-new referents that do co-occur with gestures (52%). This means that there must be other aspects (possibly related to information status) which affect the presence of gestures in general, and with first mentions in particular. One aspect concerns the operationalization of inferability. In the current study we only considered inferential relations between first mentioned and already-mentioned trigger entities. Previous research, however, suggests that a first mentioned entity can also be inferentially related to a previously mentioned *activity*, *time*, or *place* (see e.g., Ward and Hirschberg, 1985; Ward and Prince, 1991; Ward and Birner, 2001). For instance, after having talked about a baking situation, a speaker might refer to the referent "spoon" with a definite nominal because she considers it inferable given that people often use spoons when baking. It is worth considering such relations in future studies.

A further aspect is more linguistic in nature. Firbas (1971), in his original work on CD in discourse, suggests that the amount of CD a speech unit carries (whether it is a referential expression, a verb or any other unit of meaning) does not solely depend on information status but also on the semantics and the word order used in a given utterance. It is therefore possible to complement an analysis of information status of first mentions with, for instance, the semantics of the verbs used to introduce an entity into discourse or the position of the referent in the utterance. It is already known that semantics plays an important role in the way that gestures represent information (e.g., McNeill, 1992, 2005; Kita and Özyürek, 2003; Kendon, 2004; Gullberg et al., 2008; Gullberg, 2009, 2011; Debreslioska and Gullberg, 2020). However, it is rather unclear whether and if so how the semantics of a referential expression and/or the verb used to introduce a referent would also affect the incidence of gestures. Other studies suggest a relationship between the way speakers package information morpho-syntactically and the way that gestures represent information (e.g., Kita and Özyürek, 2003; Özyürek et al., 2005; Kita et al., 2007; Gullberg et al., 2008). However, also for these studies, it is unclear how morpho-syntactic packaging would influence the incidence of gesture rather than the mode of representation in gesture. Thus, examining the interplay between semantics, word order, and information status in discourse might provide further useful insights into why some entities occur with gestures and others do not and on the relationship between gestures and speech on the discourse level more generally.

Finally, there are other non-discursive aspects to consider. For instance, some entity properties may be particularly conducive to gesture production. Different objects afford action on them to different degrees, which in turn may affect how likely people are to gesture about them. For example, Chu and Kita (2016) found that speakers produced speech-associated gestures more often when the stimulus objects they saw afforded action (i.e., objects with a smooth surface) than when they did not (i.e., objects with a spiky surface). Another issue is familiarity. For instance, if someone is not, or supposes the addressee is not, familiar with a certain entity or action, such as decorating a cake with an icing bag, they might be more likely to gesture about it (cf. Campisi and Özyürek, 2013). Lastly, of course, it is also possible that the specific task in this study might have influenced why speakers did or did not gesture about entities. For instance, we encouraged speakers to say something about each picture, which might have led them to talk about aspects of the stories that they would have left out otherwise. When speakers leave out information in a narrative context, it is typically because the information is not relevant to the story at hand or because the information is old/given. It is therefore possible that this is the reason why some speakers refrained from gesturing about certain entities they talked about. These suggestions will have to be explored in future studies. In particular, it would be desirable to design experiments which can tease apart the different levels that seem to influence the distribution of gestures (discursive and non-discursive).

In conclusion, the study has provided new evidence that the incidence of gestures in discourse is related to the referential status of entities. The focus on *first mentions* in relation to gesture is novel and, unlike previous studies on this topic suggesting a parallel link between the modalities, this study reveals a complementary function of speech and gestures in discourse. Specifically, gestures are shown to accompany first mentioned inferable referents, which are hearer new, but discourse old, more often than first mentioned brand-new referents, which are hearer new and discourse new. We propose that speakers use gestures to signal that inferable referents, despite their inferential link to the previous discourse, are hearer new and that, consequently, addressees need to add them as new to their discourse representation. Gestures may help them do this. The findings are in line with the view that gestures and speech work together to build a coherent piece of discourse, but they further highlight the many and flexible functions that gestures can fulfill in relation to speech in general and in bimodal discourse reference in particular.

DATA AVAILABILITY STATEMENT

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

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. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Both authors were involved in the conception of the work and the interpretation of the data. SD wrote the first draft of the paper and both authors were involved in the critical revision of the paper. Both authors approve the publication of the paper and agree to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

We would like to thank Katrin Lindner and Nicole Weidinger for hosting the data collection at Ludwig-Maximilian University, Munich, Germany. We are also very grateful to Sabine Gosselke-Berthelsen and Irene Lami for reliability coding. Finally, we would like to gratefully acknowledge support from the Lund University Humanities Lab, Lund University, Sweden.

SUPPLEMENTARY MATERIAL

The dataset generated for this study can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01935/full#supplementary-material

REFERENCES

Allen, S. E. M. (2008). "Interacting pragmatic influences on children's argument realization" in *Crosslinguistic perspectives on argument structure: Implications for learnability*. eds. M. Bowerman and P. Brown (Mahwah, NJ: Lawrence Erlbaum Associates), 191–210.

Allen, S., and Schroder, H. (2003). "Preferred argument structure in early Inuktitut spontaneous speech data" in *Preferred argument structure: Grammar and architecture for function*. eds. J. W. Du Bois, L. Kumpf and W. Ashby (Amsterdam: John Benjamins), 301–338.

Ariel, M. (1988). Referring and accessibility. *J. Linguist.* 24, 65–87. doi: 10.1017/S0022226700011567

Ariel, M. (1991). The function of accessibility in a theory of grammar. *J. Pragmat.* 16, 443–463. doi: 10.1016/0378-2166(91)90136-L

Ariel, M. (1996). "Referring expressions and the +/− coreference distinction" in *Reference and referent accessibility*. eds. T. Fretheim and J. Gundel (Amsterdam, The Netherlands: John Benjamins), 13–33.

Arnold, J. E. (1998). *Reference form and discourse patterns (unpublished doctoral dissertation)*. Stanford, CA: Stanford University.

Arnold, J. E. (2008). Reference production: production-internal and addressee-oriented processes. *Lang. Cogn. Process.* 23, 495–527. doi: 10.1080/01690960801920099

Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Lang. Ling. Compass* 4, 187–203. doi: 10.1111/j.1749-818X.2010.00193.x

Birner, B. J. (2013). "Discourse functions at the periphery: noncanonical word order in English" in *Proceedings of the dislocated elements workshop (ZAS papers in linguistics 35)*. eds. B. Shaer, W. Frey and C. Maienborn (Berlin: ZAS), 41–62.

Birner, B. J., and Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam, The Netherlands: Benjamins.

Campisi, E., and Özyürek, A. (2013). Iconicity as a communicative strategy: recipient design in multimodal demonstrations for adults and children. *J. Pragmat.* 47, 14–27. doi: 10.1016/j.pragma.2012.12.007

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: University of Chicago Press.

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910. doi: 10.1037/0022-3514.76.6.893

Chu, M., and Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 257–270. doi: 10.1037/xlm0000168

Clancy, P. M. (1993). "Preferred argument structure in Korean acquisition" in *Proceedings of the 25th annual Child Language Research Forum*. ed. E. V. Clark (Stanford, CA: Centre for the Study of Language Information), 307–314.

Clark, H. H. (1975). "Bridging" in *Proceedings of the 1975 workshop on theoretical issues in natural language processing*. Association for computational linguistics, 169–174.

Clark, H. H. (1977). "Inferences in comprehension" in *Basic processes in reading: Perception and comprehension*. eds. D. LaBerge and S. J. Samuels (Hillsdale, NJ: Lawrence Erlbaum Associates), 243–263.

Debreslioska, S., and Gullberg, M. (2019). Discourse reference is bimodal: how information status in speech interacts with presence and viewpoint of gestures. *Discourse Process.* 56, 41–60. doi: 10.1080/0163853X.2017.1351909

Debreslioska, S., and Gullberg, M. (2020). The semantic content of gestures varies with definiteness, information status and clause structure. *J. Pragmat.* 168, 36–52. doi: 10.1016/j.pragma.2020.06.005

Debreslioska, S., Özyürek, A., Gullberg, M., and Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Process.* 50, 431–456. doi: 10.1080/0163853X.2013.824286

Fillmore, C. J. (1982). "Frame semantics" in *Cognitive linguistics: Basic readings*. ed. D. Geeraerts (Berlin, Germany: De Gruyter Mouton), 373–400.

Firbas, J. (1971). On the concept of communicative dynamism in the theory of functional sentence perspective. *Brno Studies in English*, *Vol. 7*. Brno University, Brno, Czechoslovakia, 12–47.

Foraker, S. (2011). "Gesture and discourse: how we use our hands to introduce and refer back" in *Integrating gestures: The interdisciplinary nature of gesture*. eds. G. Stam, M. Ishino and R. Ashley (Amsterdam, The Netherlands: Benjamins), 279–292.

Fraurud, K. (1990). Definiteness and the processing of noun phrases in natural discourse. *J. Semant.* 7, 395–433. doi: 10.1093/jos/7.4.395

Givón, T. (ed.) (1983). "Topic continuity in discourse: an introduction" in *Topic continuity in discourse: A quantitative cross-language study* (Amsterdam, The Netherlands: John Benjamins), 1–42.

Givón, T. (1995). "Coherence in text vs. coherence in mind" in *Coherence in spontaneous text*. eds. M. A. Gernsbacher and T. Givón, (Amsterdam, The Netherlands: John Benjamins), 59–115.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish*. Lund: Lund University Press.

Gullberg, M. (2003). "Gestures, referents, and anaphoric linkage in learner varieties" in *Information structure, linguistic structure and the dynamics of language acquisition*. eds. C. Dimroth and M. Starren (Amsterdam, The Netherlands: Benjamins), 311–328.

Gullberg, M. (2006). Handling discourse: gestures, reference tracking, and communication strategies in early L2. *Lang. Learn.* 56, 155–196. doi: 10.1111/j.0023-8333.2006.00344.x

Gullberg, M. (2009). Reconstructing verb meaning in a second language. How English speakers of L2 Dutch talk and gesture about placement. *Annu. Rev. Cogn. Linguist.* 7, 222–245. doi: 10.1075/arcl.7.09gul

Gullberg, M. (2011). "Language-specific encoding of placement events in gestures" in *Event representation in language and cognition*. eds. J. Bohnemeyer and E. Pederson (Cambridge, UK: Cambridge University Press), 166–188.

Gullberg, M., Hendriks, H., and Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Lang.* 28, 200–236. doi: 10.1177/0142723707088074

Gullberg, M., and Indefrey, P. (2003). *Language background questionnaire.* Nijmegen: Max Planck Institute for Psycholinguistics.

Gundel, J. K. (1996). "Relevance theory meets the givenness hierarchy: an account of inferrables" in *Reference and referent accessibility*. eds. T. Fretheim and J. Gundel (Amsterdam, The Netherlands: John Benjamins), 141–153.

Gundel, J. K., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307. doi: 10.2307/416535

Hickmann, M., and Hendriks, H. (1999). Cohesion and anaphora in children's narratives: a comparison of English, French, German, and Mandarin Chinese. *J. Child Lang.* 26, 419–452. doi: 10.1017/S0305000999003785

Hickmann, M., Hendriks, H., Roland, F., and Liang, J. (1996). The marking of new information in children's narratives: a comparison of English, French, German and Mandarin Chinese. *J. Child Lang.* 23, 591–619. doi: 10.1017/S0305000900008965

Holler, J., and Beattie, G. (2003). How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process? *Semiotica* 146, 81–116. doi: 10.1515/semi.2003.083

Kendon, A. (1972). "Some relationships between body motion and speech" in *Studies in dyadic communication*. eds. A. Seigman and B. Pope (Elmsford, New York: Pergamon Press), 177–216.

Kendon, A. (1973). "The role of visible behaviour in the organization of face-to-face interaction" in *Social communication and movement: Studies of interaction and expression in man and chimpanzee*. eds. M. Von Cranach and I. Vine (London: Academic Press), 29–74.

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.

Kendon, A. (1986). Some reasons for studying gesture. *Semiotica* 62, 3–28. doi: 10.1515/semi.1986.62.1-2.3

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, A. (2014). "The 'poly-modalic' nature of utterances and its implication" in *The social origins of language*. eds. D. Dor, C. Knight and D. Lewis (Oxford: Oxford University Press), 67–76.

Kita, S. (1990). *The temporal relationship between gesture and speech: A study of Japanese-English bilinguals*. Chicago, IL: Department of Psychology, University of Chicago.

Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., and Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: implications for a model of speech and gesture production. *Lang. Cogn. Process.* 22, 1212–1236. doi: 10.1080/01690960701461426

Kuznetsova, A., Brockhoff, P., and Christensen, R. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Levy, E. T., and Fowler, C. A. (2000). "The role of gestures and other graded language forms in the grounding of reference" in *Language and gesture*. ed. D. McNeill (Cambridge, UK: Cambridge University Press), 215–234.

Levy, E. T., and McNeill, D. (1992). Speech, gesture, and discourse. *Discourse Process.* 15, 277–301. doi: 10.1080/01638539209544813

Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. Los Angeles, CA: Sage.

Marslen-Wilson, W. D., Levy, E., and Komisarjevsky Tyler, L. (1982). "Producing interpretable discourse: the establishment and maintenance of reference" in *Language, place, and action: Studies in deixis and related topics*. eds. R. J. Jarvella and W. Klein (Chichester, UK: Wiley), 339–378.

McNeill, D. (1992). *Hand and mind*. Chicago, IL: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.

McNeill, D., and Levy, E. T. (1993). Cohesion and gesture. *Discourse Process.* 16, 363–386. doi: 10.1080/01638539309544845

Narasimhan, B., Budwig, N., and Murty, L. (2005). Argument realization in Hindi caregiver–child discourse. *J. Pragmat.* 37, 461–495. doi: 10.1016/j.pragma.2004.01.005

Özyürek, A., Kita, S., Allen, S., Furman, R., and Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture* 5, 219–240. doi: 10.1075/gest.5.1-2.15ozy

Parrill, F. (2012). "Interactions between discourse status and viewpoint in co-speech gesture" in *Viewpoint in language: A multimodal perspective*. eds. B. Dancygier and E. Sweetser (Cambridge, UK: Cambridge University Press), 97–112.

Perniss, P., and Özyürek, A. (2015). Visible cohesion: a comparison of reference tracking in sign, speech, and co-speech gesture. *Top. Cogn. Sci.* 7, 36–60. doi: 10.1111/tops.12122

Prince, E. F. (1981). "Toward a taxonomy of given-new information" in *Radical pragmatics*. ed. P. Cole (New York: Academic Press), 223–256.

Prince, E. F. (1992). "The ZPG letter: subjects, definiteness and information status" in *Discourse description: Diverse analyses of a fund raising text*. eds. S. Thompson and W. Mann (Amsterdam, The Netherlands: Benjamins), 295–325.

RStudio Team (2016). RStudio: Integrated development for R. RStudio, Inc., Boston, MA. Available at: http://www.rstudio.com/

Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Appl. Psycholinguist.* 26, 437–462. doi: 10.1017/S0142716405050241

Seyfeddinipur, M. (2006). Disfluency: *Interrupting speech and gesture (MPI Series in Psycholinguistics)*. [Unpublished doctoral dissertation]. Nijmegen, The Netherlands: Radboud University Nijmegen.

Sloetjes, H., and Wittenburg, P. (2008). "Annotation by category – ELAN and ISO DCR" in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*; May 28–June 1, 2008; Marrakech, Morocco.

So, W. C., Demir, Ö. E., and Goldin-Meadow, S. (2010). When speech is ambiguous, gesture steps in: sensitivity to discourse-pragmatic principles in early childhood. *Appl. Psycholinguist.* 31, 209–224. doi: 10.1017/S0142716409990221

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *Int. Rev. Appl. Ling.* 44, 143–169. doi: 10.1515/IRAL.2006.006

Ward, G., and Birner, B. J. (2001). "Discourse and information structure" in *Handbook of discourse analysis*. eds. D. Schiffrin, D. Tannen and H. Hamilton (Oxford: Basil Blackwell), 119–137.

Ward, G., and Hirschberg, J. (1985). Implicating uncertainty: the pragmatics of fall-rise intonation. *Language* 61, 747–776. doi: 10.2307/414489

Ward, G., and Prince, E. F. (1991). On the topicalization of indefinite NPs. *J. Pragmat.* 16, 167–177. doi: 10.1016/0378-2166(91)90079-D

Wilkin, K., and Holler, J. (2011). "Speakers' use of 'action' and 'entity' gestures with definite and indefinite references" in *Integrating gestures: The interdisciplinary nature of gesture*. eds. G. Stam and M. Ishino (Amsterdam, The Netherlands: Benjamins), 293–308.

Wu, S. (2018). *Multimodality of constructions in construction grammars: Transitivity, transitivity alternations, and the dative alternation (doctoral dissertation)*. Amsterdam, The Netherlands: Free University Amsterdam.

Yoshioka, K. (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse (doctoral dissertation)*. Groningen, The Netherlands: University of Groningen.

Yoshioka, K. (2008). Gesture and information structure in first and second language. *Gesture* 8, 236–255. doi: 10.1075/gest.8.2.07yos

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

frontiers
in Psychology

ORIGINAL RESEARCH
published: 19 October 2020
doi: 10.3389/fpsyg.2020.575929

Beat Gestures for Comprehension and Recall: Differential Effects of Language Learners and Native Listeners

***Patrick Louis Rohrer*[1,2]*, *Elisabeth Delais-Roussarie*[1] *and Pilar Prieto*[2,3]**

[1] Université de Nantes, UMR 6310, Laboratoire de Linguistique de Nantes (LLING), Nantes, France, [2] Grup d'Estudis de Prosòdia, Department of Translation and Language Sciences, Pompeu Fabra University, Barcelona, Spain, [3] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Previous work has shown how native listeners benefit from observing iconic gestures during speech comprehension tasks of both degraded and non-degraded speech. By contrast, effects of the use of gestures in non-native listener populations are less clear and studies have mostly involved iconic gestures. The current study aims to complement these findings by testing the potential beneficial effects of beat gestures (non-referential gestures which are often used for information- and discourse marking) on language recall and discourse comprehension using a narrative-drawing task carried out by native and non-native listeners. Using a within-subject design, 51 French intermediate learners of English participated in a narrative-drawing task. Each participant was assigned 8 videos to watch, where a native speaker describes the events of a short comic strip. Videos were presented in random order, in four conditions: in Native listening conditions with frequent, naturally-modeled beat gestures, in Native listening conditions without any gesture, in Non-native listening conditions with frequent, naturally-modeled beat gestures, and in Non-native listening conditions without any gesture. Participants watched each video twice and then immediately recreated the comic strip through their own drawings. Participants' drawings were then evaluated for discourse comprehension (via their ability to convey the main goals of the narrative through their drawings) and recall (via the number of gesturally-marked elements in the narration that were included in their drawings). Results showed that for native listeners, beat gestures had no significant effect on either recall or comprehension. In non-native speech, however, beat gestures led to significantly lower comprehension and recall scores. These results suggest that frequent, naturally-modeled beat gestures in longer discourses may increase cognitive load for language learners, resulting in negative effects on both memory and language understanding. These findings add to the growing body of literature that suggests that gesture benefits are not a "one-size-fits-all" solution, but rather may be contingent on factors such as language proficiency and gesture rate, particularly in that whenever beat gestures are repeatedly used in discourse, they inherently lose their saliency as markers of important information.

Keywords: gesture, comprehension, recall, beat gestures, non-referential gestures, L1/L2

OPEN ACCESS

Edited by:
Naomi Sweller,
Macquarie University, Australia

Reviewed by:
Linda Drijvers,
Radboud University, Netherlands
Marieke Hoetjes,
Radboud University Nijmegen,
Netherlands

****Correspondence:***
Patrick Louis Rohrer
patrick.rohrer@upf.edu

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

Received: *24 June 2020*
Accepted: *28 September 2020*
Published: *19 October 2020*

Citation:
Rohrer PL, Delais-Roussarie E and Prieto P (2020) Beat Gestures for Comprehension and Recall: Differential Effects of Language Learners and Native Listeners. Front. Psychol. 11:575929. doi: 10.3389/fpsyg.2020.575929

INTRODUCTION

Speech is a multimodal act that allows for listeners to make use of both auditory as well as visual cues to make sense of the incoming message. Numerous studies have shown that speech produced with referential gestures[1] boost both comprehension and recall in the L1 (Riseborough, 1981; Cohen and Otterbein, 1992; among many others), with very few studies showing no effects (e.g., Austin and Sweller, 2014; Dahl and Ludvigsen, 2014). Similarly, positive results have also been found in the L2 (Sueyoshi and Hardison, 2005; Tellier, 2008; Kelly et al., 2009; Macedonia et al., 2011; among many others). A meta-analysis by Hostetter (2011) which analyzed over 60 studies describes six ways in which referential gestures may boost memory, comprehension, and learning: (i) By being better adapted at conveying spatial information than speech, (ii) by giving additional information that is not in speech, (iii) by having positive effects on the speaker's speech production, (iv) by presenting information that is redundant with speech, affording listeners additional cues to glean meaning, (v) by capturing a listener's attention, and (vi) by boosting a positive rapport between speaker and listener. Further evidence of these beneficial effects is found in electrophysiological studies on the semantic integration of referential gestures. A handful of studies have found that iconic gestures that are incongruent with their lexical referent in speech produce large N400 s, indicating difficulty in integrating semantic meaning (e.g., Holle and Gunter, 2007; Kelly et al., 2010 among others).

What is less well understood, however, is under which conditions iconic gestures benefit recall and comprehension processes the most. For example, a recent study by Dargue and Sweller (2020) found that *typically* produced iconic gestures aided comprehension of a short narrative over *atypically* produced iconic gestures (e.g., moving one hand upward while pointing to the ceiling with the other hand to represent the character picking up a bucket). Similarly, electrophysiological studies have also determined that N400 effects can be modulated by factors such as speaker style (i.e., using only iconic gestures, compared to producing iconic gestures along with meaningless grooming movements; Obermeier et al., 2015), the temporal affiliation between the iconic gesture and its lexical referent (Obermeier et al., 2011), noise conditions (Drijvers and Özyürek, 2017), or native-language status (Ibáñez et al., 2010; Drijvers and Özyürek, 2018). The current study aims to deepen our knowledge regarding the factor of native-listener status.

Indeed, an important speaker-external factor that seems to strongly regulate the effectiveness of gesture is native-language status. When directly comparing the effect of gestures on recall and comprehension by native and non-native listeners, a different pattern of results emerges depending on L2 proficiency level. Following previous EEG studies with a gesture-congruency paradigm with referential gestures, Ibáñez et al. (2010) found that while high-proficiency learners of German showed similar patterns to native listeners in N400 modulation, low-proficiency learners showed no modulation. The interpretation of these results was such that when speakers are at a lower proficiency, they do not even attempt to integrate information in gesture. Along those same lines, Drijvers and Özyürek (2018) found that iconic gestures in clear speech conditions resulted in larger N400 components in intermediate-level non-native listeners than native listeners, while no N400 modulation was found for non-native listeners in degraded speech. The authors interpret this larger N400 effect in non-native listeners as evidence that they need to focus more strongly on gestures in clear speech to integrate the semantic information. However, when there are no phonological cues available to help with the process, they no longer make use of gestures for semantic integration.

In a recent eye-tracking study, Drijvers et al. (2019a) expanded upon these results. The authors presented native and highly proficient non-native listeners a set of Dutch verbs that were uttered either with or without gesture, and in clear and degraded speech. Immediately following the presentation of each video stimulus, participants were asked to choose which word they heard from four potential candidates. Even though the results showed that both native and non-native speakers benefited from the presence of gesture for the comprehension of Dutch words produced in isolation, they crucially found that both native and non-native listeners showed more accurate answers and faster response times in the gesture condition than in the no gesture condition. While language status did not affect the accuracy of responses, native listeners answered quicker than non-native listeners in the gesture condition with degraded speech. The eye-tracking data showed that in the gesture condition with degraded speech, while all listeners focused more on the face than the gesture, non-native listeners tended to fixate more on gestures than native listeners. Thus authors suggest that non-native listeners cannot make use of visual information from the mouth when auditory cues are unavailable, and thus look for visual information elsewhere. This is unlike native speakers, who can indeed make use of visual cues from the mouth and integrate information both from manual and mouth movements simultaneously. It is this efficiency in integrating multiple channels of information simultaneously that leads the native speakers to respond faster in the cued recall task described above (see also Drijvers et al., 2019b).

To our knowledge, only one study has assessed the benefits of the presence of iconic gestures on recall and comprehension by native and non-native listeners using larger discourses. Dahl and Ludvigsen (2014) directly compared the effects of iconic gestures in both native and non-native listeners in terms of recall and comprehension in a cartoon picture drawing task. 28 native English speaking adolescents and 46 Norwegian adolescents who had been learning English for 7–8 years participated in the study. Each group of participants were divided into two experimental conditions, resulting in a total of 4 experimental groups: Native listener with gesture (NL-G), Native listener without gesture (NL-NonG), Foreign listener with gesture (FL-G), and Foreign listener without gesture (FL-NonG). Each group saw the same 4 picture descriptions presented in English, differing only in whether referential gestures were present or not. Upon watching each video, participants were asked to reproduce the picture that

[1] Studies on co-speech gestures have widely adopted McNeill's (1992) classification of gestures as iconic, metaphoric, deictic, or beat gestures. The former three are considered referential in nature, as they make direct references to semantic content in speech.

had just been described. Their drawings were evaluated in terms of explicit recall (the presence of elements explicitly described in the discourse), implicit comprehension (the presence of logically implied elements), distortions (elements that were present but inaccurately portrayed), and based on these measures, a composite score was calculated. They found that the native language groups performed similarly on the task regardless of the presence or absence of gesture. In the FL groups, however, the G group showed much higher scores of recall and comprehension, and fewer distortions than their NonG counterparts. Indeed, the FL-G group performed just as well as both NL groups. These results suggest that referential gestures may not have an effect in native listeners, while non-native listeners benefit from information coded in gesture.

Importantly, compared to their referential counterparts, fewer studies have investigated effects on comprehension and recall when there is no lexico-semantic meaning associated with the gesture[2]. Indeed, non-referential beat gestures are one of the most common types of gesture that are produced by speakers, particularly the case in academic contexts where these gestures predominate at rates of up to 94.6% of the gesture types produced (Shattuck-Hufnagel and Ren, 2018, see also Rohrer et al., 2019 for similar results). These gestures (much like their referential counterparts) are also integrated with speech prosody (often co-occurring with pitch accentuation), and their presence can actually modulate a listener's perception of prominence (see Krahmer and Swerts, 2007; Bosker and Peeters, 2020). Further, non-referential beat gestures have important discursive and pragmatic functions, such as marking information structure (Im and Baumann, 2020), epistemic stance (Prieto et al., 2018; Shattuck-Hufnagel and Prieto, 2019), among others. Indeed, these gestures work with prosodic prominence to act as "highlighters" to important information in speech, potentially increasing listeners' attention to key words in speech. Thus it seems important to understand how these movements are processed by listeners and can potentially aid in discourse comprehension and recall. This is especially true in the case of non-native listeners, as these movements may aid in determining important aspects of speech and boosting comprehension, particularly in the language classroom. Conversely, they may also be a distraction from concentrating on decoding speech in the auditory domain, due to their non-imagistic nature. To our knowledge no study has assessed the effects of beat gestures on comprehension and recall by non-native listeners. The current study investigates for the first time the potential beneficial effects of beat gestures on language recall and comprehension of a narrative task by both native and non-native listeners.

Recent electrophysiological evidence has helped in obtaining more insight on the integration of non-referential gestures with speech, revealing that non-referential beat gestures boost attention and can help ease semantic integration. An early study by Biau and Soto-Faraco (2013) found that beat gesture-accompanied words elicited a positive shift in the early stages of processing, as well as a later positivity around 200 ms after word onset, showing that gesture is integrated early on in speech processing. Similarly, a study by Dimitrova et al. (2016) found that beat gestures elicited a positivity around 300 ms after word onset. They attribute this to a "novel P3a" component that is said to reflect increased attention. These two studies, when taken together, support the idea of beat gestures working as a "speech highlighter," boosting attention. Another study by Wang and Chu (2013) showed that beat gestures elicited smaller N400 components, independently of pitch accentuation. Thus, the authors conclude that beat gestures attract attention to focused words, ultimately facilitating their semantic integration. However, while electrophysiological studies seem to suggest that non-referential beat gestures boost attention and ease semantic integration, behavioral studies on these gestures have found conflicting results on their effects on recall and comprehension patterns.

Despite the aforementioned electrophysiological results, behavioral studies have found mixed results when assessing the use of non-referential beat gestures on recall and comprehension patterns, both in adults and children. Comparing gesture types, Feyereisen (2006) exposed adults to 26 sentences, where 10 sentences contained a referential gesture, 10 contained a non-referential gesture, and 6 were filler sentences. A free-recall task showed that the participants remembered the sentences with referential gestures more than those with non-referential gestures. On the other hand, So et al. (2012) found that when presenting lists of single words accompanied by either iconic, beat, or no gesture, adults benefited equally from both iconic and beat gestures, while children only benefited from iconic gesture. However, the previous two studies presented sentences and words without any context. Again looking at both adults and children, Austin and Sweller (2014) investigated the effects of different gesture types on the recall of spatial directions. Participants were shown a Lego base plate with arranged Lego pieces representing different destinations in a town. Participants were then told by the researcher the path the Lego man took. The researcher described the path in one of three conditions: (a) no gesture, (b) producing 20 beat gestures, or (c) producing a combination of gestures (iconic, deictic, metaphoric, and beat gestures, $N = 5$ per type). After hearing the spatial direction describing the Lego man's path and a 120 s join-the-dots filler activity, participants were asked to recount the path that was described to them. Contrary to the results from So et al. (2012), they found that children did benefit from both "combined" gesture and beat gesture conditions, while adults did not show any beneficial effects from either gesture condition. Further studies with children have shown mixed results. While studies like Igualada et al. (2017) and Llanes-Coromina et al. (2018) found beneficial effects of beat gestures in lists and short discourse contexts with one target beat gesture per sentence, Macoun and Sweller (2016) found that there was no benefit from the presence of beat gesture produced in larger narrations describing a girl's afternoon in the park with her family. When comparing the effects of non-referential

[2] i.e., McNeill's "beat" gesture, often said to be non-referential because there is no clear semantic reference in speech. Indeed, [McNeill (1992):15] describes these movements as simple up-and-down movements or flicks of the hand or fingers, that seem to be beating to the rhythm of speech. Recently, they have been claimed to have more pragmatic functions. For example, these gestures tend to mark new or contrastive information, or discourse structure (see also Prieto et al., 2018; Shattuck-Hufnagel and Prieto, 2019).

beat gestures, most studies have justified their disparate results by focusing on methodological differences, particularly in terms of stimuli presentation patterns. Some studies presented single words or sentences out of context, while others offered longer narratives of varying sizes. It is important to note that the studies on children seem to suggest that beat gestures are most effective when marking focused information in a pragmatically relevant context. While Igualada et al. (2017) and Llanes-Coromina et al. (2018) used short discourses or lists of words with one gesture occurring in a pragmatically relevant position, studies by Austin and Sweller (2014) and Macoun and Sweller (2016) used a more difficult task with a 2-min narrative with a larger occurrence of beat gestures marking the same words as in the referential gesture condition [20 gestures within 10 target sentences for the Beat gesture condition in Austin and Sweller (2014); 10 gestures within a 2 min narrative for the Beat gesture condition in Macoun and Sweller, 2016]. In this context we think that it is especially relevant to assess the effects of beat gestures in natural speech conditions, which may contain multiple gestures within one narration.

Two studies with adults complemented the data obtained with children, and took into account the relationship between beat gestures and prosody. They showed that gestures are most effective when coupling with prosody to mark contrastively focused information in a pragmatically relevant context. Kushch and Prieto (2016) used larger discourses that contained two contrastive sets within the narrative. The discourses were produced so that prominence could either be given prosodically (through L + H* pitch accentuation) or prosodically and gesturally (with both a pitch accent and a non-referential beat gesture). These conditions could either appear on the first contrastive pair (where the second pair would be unaccented) or vice versa, resulting in four possible configurations. 20 native Catalan speaking participants watched the discourses and were subsequently given a cloze task, where they had to fill in the words that were contrastively focused from each pair. They found that beat gestures boosted recall significantly more than prosodic prominence alone, and that this effect was even greater when it accompanied the first contrastive pair in discourse. These results were further refined in a more recent study by Morett and Fraundorf (2019). Using similar discourses, they manipulated the conditions to have beat gesture present or absent, and accenting be either presentational (H*) or contrastive (L + H*). While they did not find a main effect of gesture on the recall of information, they did find that contrastively marked information accompanied by a beat gesture was remembered more than presentational information when it was marked with a beat gesture. When gestures were absent, there was no effect of pitch accent type. In other words, beat gestures seem to modulate the efficacy of contrastively marked prominence. Thus, these studies suggest that the gesture's pragmatic function is also a factor that affects beat gesture's efficiency in boosting recall and comprehension.

All in all, there is a clear need to assess why non-referential beat gestures seem to have a positive impact on language processing in some instances but not in others. In this regard, following up on recent studies focusing on referential gestures, some research has begun investigating *under which conditions* beat gestures are helpful. To our knowledge, only three studies have assessed the role of beat gestures for non-native listeners, particularly focusing on their effects in novel vocabulary learning, with mixed results. Levantinou and Navarretta (2015) followed the same methodology as So et al. (2012) with presenting individual words with or without iconic gesture, beat gesture, or no gesture. They found that only iconic gestures boosted recall, and that there was no significant difference between the beat and no gesture conditions. The authors claimed that beat gestures may have in fact increased the learners' cognitive load, as they have not yet learned how to interpret these gestures. Another study by Kushch et al. (2018) presented novel Russian vocabulary words to naïve Catalan learners in a carrier sentence, such as "Bossa es diu 'sumka' en rus" (translation: "*Bag is called 'sumka' in Russian*"). The target word (*sumka*) was presented in 4 conditions: Accompaniment with neither a (L + H*) pitch accent nor a gesture; Accompanied with a (L + H*) pitch accent (no gesture); Accompanied with a gesture (no pitch accent); or Accompanied with both a (L + H*) pitch accent and gesture. They found that the participants recalled best when target words were produced with both a gesture and a pitch accent. When only one prominence was produced, pitch accented words were better remembered than words produced with beat gesture only. The authors thus claimed that beat gestures can be beneficial in restricted learning contexts and when they co-occur with focal pitch accents. Finally, a study by Morett (2014) used an interactive word teaching and learning task with pairs of native English speakers with no knowledge of Hungarian to assess gesture's effect on the recall of novel vocabulary. For each pair, one participant was designated as the "explainer" and the other as the "learner." The explainer had to teach a total of 20 novel Hungarian words. After the presentation of each word, the explainer had to teach the learner the novel vocabulary word "however, they thought [the learner] would learn it best" (i.e., they had no specific instructions regarding gesture production). The entire interaction between the two participants was filmed. After the filmed learning phase, participants had to take a recall test. Gesture's impact was determined by using multiple regression analysis to examine the relationship between gesture production by both participants during the learning phase and their recall scores. They found that observing gesture did not predict word recall for either participant, regardless of type. However, explainers' production of beat gestures did predict their own word recall, while learners' representational gesture production predicted their own word recall. The author explains that these divergent results may be due to the fact that learners may have used representational gesture to enrich the conceptual links between the new words and their referents, while the explainers may have made use of reinforced verbal associations that were established while using beat gestures to convey the meaning of target words. The overall results from this study suggest that gesture production is more beneficial than their mere perception, and in regards to beat gestures, they may be beneficial for different speakers in different contexts. Thus, studies involving the use of beat gesture in L2 have found conflicting results. Further, none of these studies have directly compared native listeners to non-native listeners in the recall and comprehension of complex discourses.

In sum, the previous research on the effects of non-referential beat gestures for recall and comprehension has shown mixed results, where positive results have generally been shown when beat gestures are used in pragmatically restricted contexts, e.g., marking contrastively focused information. Less is known regarding the effects of beat gesture production that has been modeled after natural discursive speech, reflecting more natural, real-world experiences that listeners encounter (yet see Austin and Sweller, 2014; Macoun and Sweller, 2016). Thus it seems important to see the effects of these gestures in more natural speech conditions, which may contain multiple gestures within one narration. Importantly, no study with beat gestures has directly compared between native and non-native listeners. Thus the main aim of the study is to compare the effects of beat gestures between native and low-intermediate-level non-native listeners in a narrative-drawing task. This population was chosen as some studies have suggested that gestures may be more beneficial for lower-level learners (see Sueyoshi and Hardison, 2005; Morett, 2014). We believe that non-referential beat gestures may help non-native listeners as they index key words in the narrative, potentially boosting their attention to these aspects and consequently aiding in their recall. Further, as mentioned before, beat gestures serve discourse and information structure marking functions, which may boost discourse comprehension in terms of understanding the relationship between the elements and actions in the narrative. However, it is quite possible that compared to native speakers, these more complex, naturalistic conditions may lead to cognitive overload (i.e., processing costs beyond the listener's cognitive capacity) for low-intermediate-level non-native listeners with too much visual stimulation, causing them to focus on the movements and miss out on important information being presented orally (e.g., Drijvers et al., 2019a). Following Dahl and Ludvigsen (2014), a narrative-drawing task was chosen as it offers a blank slate to determine what information is recalled and understood from the narrative, without the implications of using comprehension questions which may assess recall and comprehension in a more precise manner but require language processing and production skills to answer. This is particularly relevant for low-intermediate L2 learners. The current study will thus give insight on the effects of non-referential gestures on recall and discourse comprehension in more natural contexts and particularly by low-intermediate non-native listeners, which could potentially guide our understanding on not just *if* these gestures are beneficial, but *when* they are beneficial. The results may also eventually be applied in language learning contexts, where gestures may be used to potentially boost vocabulary learning or facilitate oral comprehension in the L2.

MATERIALS AND METHODS

Participants

A total of 51 participants (41 females, 9 males, and 1 non-binary, $M_{\text{age}} = 23.28$, SD = 7.2) were recruited from 4 intermediate-level English classes at the University of Nantes. One of the English classes where participants were recruited from was for second year undergraduate students studying English as part of their Language Sciences degree (N = 15). The other three English classes were offered by the *Service Universitaire des Langues* (SUL) at the University of Nantes (N = 13, 13, and 10, respectively). These courses are open to all students and faculty wishing to improve their English level. Professors from each course agreed to dedicate one class session to the experiment. All participants gave informed consent.

In order to assess L2 level, participants in the 3 SUL classes had taken the University of Grenoble's SELF language assessment[3] test before enrolling in the class. Students who did not have a SELF score were given the 20-min General English Test offered by International House London[4] before the task. A large majority of participants reported an intermediate level of English (CEFR: A1 = 5%, A2 = 5%, B1 = 35%, B2 = 50%, and C1 = 5%).

Eight students were removed from analyses. First, 6 students were removed because they reported languages other than French as their L1. Two students were removed from analyses for having a C1 level in English. Since previous research has shown that advanced learners attend to gestures in much the same way as native listeners (Ibáñez et al., 2010), these participants' profiles were deemed too native-like and did not match the profile of the rest of the students.

Materials

Stimuli Creation for the Drawing Task

A subset of 8 comic strip illustrations was chosen from the Simon's Cat comic series that were used in Cravotta et al. (2019). These 8 comic strips were chosen based on the ease of translating the illustrations to a short narrative that could be understood by low-intermediate level language learners. A short narration was then written for each comic in both French and English. All narratives followed the same basic structure where each square in the comic trip was introduced by a sequencing marker ("First," "Next," "Then," and "Finally") which described the development of the narrative, followed by a short description of the orientation of items in the square or actions that have occurred since the previous square. See **Figure 1** for an example comic strip; its corresponding narration can be found in section "**Appendix A**."

Gesture placement for the final stimuli was then determined by recordings of two native speakers in each language who read the narrations aloud. The 4 speakers had no knowledge of the purpose of the study and were merely asked to read the narration while being "expressive with their hands." In doing so, it was possible to determine the most natural lexical affiliates in the narrative to be marked with a gesture. A majority of the gestures produced were non-referential in nature. The lexical affiliates of each gesture (regardless of referentiality) produced by each participant for each comic was then determined, and the inclusion of these "gesturally-marked elements" in the final stimuli were determined by three factors. First, a gesturally-marked element was included if at least 3 speakers marked

[3] Innovalangues: SELF, http://innovalangues.fr/realisations/systeme-d-evaluation-en-langues-a-visee-formative/

[4] International House London, https://www.testmylevel.com/

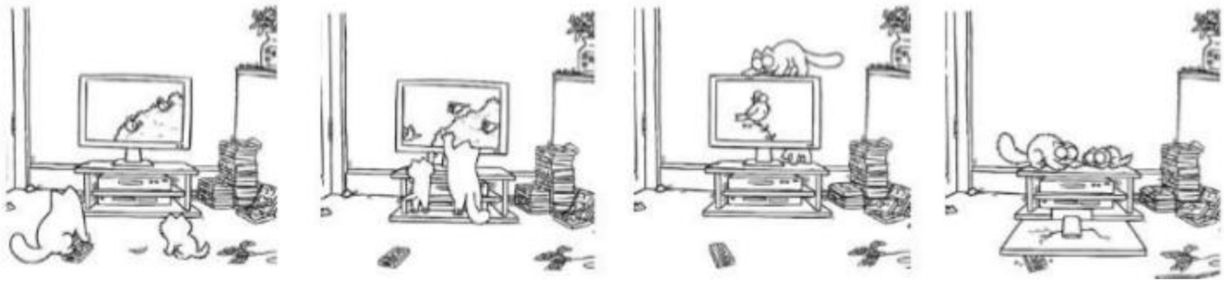


FIGURE 1 | An example comic strip, taken from "Simon's Cat" by Simon Tofield. Reprinted with permission. © Simon's Cat Ltd.

that same word-referent (across languages) was automatically included in the final stimuli. A second factor was gesture salience (i.e., the perception of a large gesture movement or more emphatic gesture). If one of the speakers made a particularly salient gesture on a word (and perhaps one other person also marked that same word with a gesture), then it was included in the final stimuli. The third and final factor was the pertinence of the gesture to the narrative. In other words, if 2 speakers marked a word that contrasted with another element, it was seen as being pertinent to the narrative as it disambiguated two items, and this gesture would be included in the final stimuli. After analyzing the natural speech productions, scripts were created for each narrative that contained the gesturally-marked lexical affiliates in bold for filming. **Table 1** shows the average number of gestures per sentence, the total number of gestures, and the duration of each video.

Video Filming, Editing and Validation of the Target Narrations for the Drawing Task

Two female native speakers were recruited to record the spoken narrations in their respective native language. Recordings took place in a professional recording studio at Universitat Pompeu Fabra in Barcelona, and the speakers were paid 10 euros per hour. The actresses were briefly shown the types of gestures they would be making (i.e., beat gesture) and that they would produce them on target words. They were then given opportunities to practice producing the narratives with gestures. While the speakers were relatively free to produce the non-referential gestures as they saw fit, they were given feedback to have a more relaxed, natural style. This was done to avoid disparate differences in gesture salience between the two speakers. In other words, both speakers were trained to produce the gestures in a relaxed and natural way, with most gestures being small up-and-down movements or flips. Each actress then recorded multiple trials of each narration in both the Gesture (G) and the No-Gesture (NG) conditions following a teleprompter which displayed a script of each narrative, with target words to contain a gesture (in the G condition) being marked in capital letters. In order to maintain a natural style, no instructions were given in terms of prosodic emphasis.

Following the recording session, videos were then edited in Adobe Premiere Pro (CS6). Videos were edited to show the actresses placed in front of a simple gray background. They were shown from the waist up so that both hands were visible, as well as the face. Again, this was done to keep the stimuli close to real-world situations as possible. The average duration of the edited videos of the target narrations was 63.94 s (±9.27 s) and each narration contained an average of 25 (±5) gestures. **Figure 2** shows four still-frames taken from one of the narrations showing each speaker either in the gesture condition, or the no-gesture condition.

To ensure the naturalness of the gestures in the video stimuli, 5 native English speakers, and 6 native French speakers evaluated how natural the gestures seemed for each video of their corresponding language. Each rater evaluated the videos on a Likert scale from 1 to 7, where 7 was the most natural. The French videos received an average score of 4.71 (SD = 1.58) while the English videos received an average score of 5.53 (SD = 1.48). This suggests that raters generally felt that the videos were relatively natural-looking.

Stimuli Organization of the Narrations for the Drawing Task

The aforementioned steps resulted in a total of 32 videos, where each narration was video-recorded in four conditions: in native listening conditions with gesture, in native listening conditions without gesture, in non-native listening conditions with gesture, and in non-native listening conditions without gesture. In order to ensure that participants see all of the narratives in different language and gesture conditions in a balanced manner, a Latin-square method allowed for the division of the stimuli into 4 stimuli lists (see **Table 2**), where narrations were balanced for the language listening condition and gesture condition in each list. In other words, each list contained the 8 narrations, but the lists differed in terms of the language and gesture conditions that were presented for each narration. By organizing the stimuli this way, it was possible to avoid any bias stemming from the individual narrations themselves. Each stimuli list was then uploaded to SurveyGizmo as an individual online survey for the presentation of the stimuli. Each survey followed the same structure. An initial screen gave instructions. The survey then alternated between video presentation screens and screens that instructed participants to draw. The survey ended with a "Thank You" screen that informed participants that the experiment had ended (see section "Drawing task" for more details).

TABLE 1 | The average number of gestures per sentence, and total number of gestures per comic narrations.

Comic language	**English**			**French**		
Comic number	**Average N of gestures per sentence**	**Total N gestures**	**Video duration (in seconds)**	**Average N of gestures per sentence**	**Total N gestures**	**Video duration (in seconds)**
1	2.75	22	64 s	2.33	21	64 s
2	1.71	24	84 s	1.92	25	75 s
3	3	21	53 s	3	21	50 s
4	2.08	27	73 s	1.75	21	62 s
5	3	30	77 s	2.91	32	73 s
6	3.09	34	65 s	2.82	31	59 s
7	2.88	23	54 s	2.3	23	53 s
8	1.91	21	61 s	2.1	21	60 s

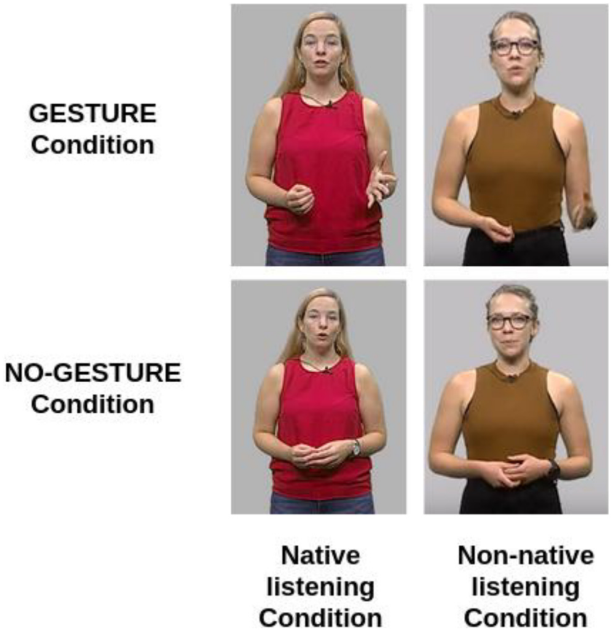


FIGURE 2 | Still-frames taken from the stimuli videos of one narration in each condition.

Procedure

An online linguistic background survey was emailed to each participant to be completed before the drawing task in order to collect each participant's personal information (e.g., gender, age, level of study), as well as to assess their L1. Participants were also asked to bring their own laptops and headphones on the day of the drawing task, which would allow them to access the survey online. Immediately before the session, each participant was given a link to their corresponding list's online survey containing the stimuli videos.

Drawing Task

The drawing task was carried out in 4 English classes containing about 15 students each. The participants did the task individually and in a self-paced manner. It was carried out in a quiet classroom under the supervision of the class instructor[5]. Each participant was given a small task booklet that contained an instructions page, followed by a set of 8 pages, where each page contained 6 large squares for the participants to draw their interpretations of the comic narrations. Then, participants were informed that they were going to perform a narrative comprehension task, and were directed to read the instructions carefully. Instructions (adapted from Dahl and Ludvigsen, 2014) were available in both French and English and were as follows:

> *You are about to watch 8 short video clips, half of them in French, and half of them in English. Each clip is a description of a different humorous comic strip. Watch to the first description and create a picture in your mind of what this comic strip looks like. Try to remember as many details as possible. You are not allowed to draw while you are watching the video. Once the video has ended, try to draw the comic strip that you just heard described.*
>
> *The quality of your drawing skills is not the most important thing. What is important is how much you remember of the comic strip that was described and that you show that through what you draw. Try to include as much as possible in the drawing. In case something is hard for you to draw or some element in the drawing seems unclear in the picture, you can write and draw arrows next to the element to clarify what it is.*
>
> *You are given a page with 6 squares to draw in. Note that you can use as many or as few of the squares as you think are appropriate for the story. That is, if you think the comic being describing is only 3 squares long, you draw the entire comic in three squares. Try to use all of the space within each square.*
>
> *Once you have finished the drawing, you may move on to the next video description.*

Upon reading the instructions, participants were directed to access the survey via the link that they had received by e-mail. The online survey again gave a more concise version of the above instructions and once the participant acknowledged they understood and were prepared, they began

[5] As three of the four classes took place at the same time, the first author was present for two of the four experimental sessions. In the two experimental sessions in which the main author was not present, participants were under the supervision of the course instructor who had been debriefed about the details of the procedure. Neither instructor indicated any difficulty or issues while running the experimental session.

TABLE 2 | The distribution of language and gesture conditions by comic narration into four counter-balanced lists.

Comic narration number	List 1	List 2	List 3	List 4
1	NON-NATIVE-G	NATIVE-NG	NATIVE-G	NON-NATIVE-NG
2	NON-NATIVE-NG	NON-NATIVE-G	NATIVE-NG	NATIVE-G
3	NATIVE-G	NON-NATIVE-NG	NON-NATIVE-G	NATIVE-NG
4	NATIVE-NG	NATIVE-G	NON-NATIVE-NG	NON-NATIVE-G
5	NON-NATIVE-G	NATIVE-NG	NATIVE-G	NON-NATIVE-NG
6	NON-NATIVE-NG	NON-NATIVE-G	NATIVE.-NG	NATIVE-G
7	NATIVE-G	NON-NATIVE-NG	NON-NATIVE-G	NATIVE-NG
8	NATIVE-NG	NATIVE-G	NON-NATIVE-NG	NON-NATIVE-G

the stimuli presentation. Stimuli from the participants' assigned list were presented in random order, and the presentation screen contained an embedded video. This screen remained accessible for at least 2 min and 30 s, just enough time to watch each video two times, while not allowing participants to watch a third time. After watching the video two times (or when the time limit was reached), the survey would proceed to a screen that instructed the students to draw the comic that had just been described in the video. There was no time limit on this screen, so participants could take the time necessary to complete their drawing. Once completed with their drawing, the participant then proceeded to the next random video stimulus. Upon completing the survey and all of the drawings for the 8 narrations, students turned their booklet into the instructor.

Scoring of the Drawing Task

To evaluate **explicit recall**, a list of all the items that were gesturally marked in the gesture condition was created for each narration (see **Table 1** for the number of gesturally marked items per narration). These lists served as checklists when determining whether these specific items were accurately remembered or not. The main author carried out all of the scoring while unaware of which condition the drawing pertained to. For each item in the checklist of a given comic description, if the element is clearly remembered and present in the drawing, a score of 2 is given. If the element was not remembered exactly as described or it is ambiguous whether the element was remembered clearly or not, a score of 1 was given. This score was used for cases in which memory of the element was distorted. When the element is not present at all in the drawing, a score of 0 was given. For example, if the narration had the sentence "The cat is sleeping on a **rectangular** rug" (bold indicates the lexical affiliate of the gesture in the G-condition), and the drawing shows a rectangular rug, the participant received 2 points. If the drawing shows a circular rug, the participant would receive only 1 point. If there is no rug in the drawing, the participant received 0 points. The maximum number of points a participant could receive per drawing ranged from 38 to 60 points depending on the number of gesturally marked items in the corresponding narration. While most gesturally marked elements were nouns, verbs, or adjectives that marked focus (e.g., "a **rectangular** rug" or "the cat **jumped** in the air."), discourse markers "First," "Next," "Then," and "Finally" were also gesturally marked elements. As such, not only were participants' recall evaluated in terms of remembering particular items or actions, but also in terms of the sequencing of events. See section "**Appendix B**" for an example scoring of recall for one comic square.

Unlike the current study that uses narratives, the study by Dahl and Ludvigsen (2014) used picture descriptions as stimuli for their student to draw, and they not only looked at explicit recall, but also "implicit comprehension." They describe implicit comprehension as the participant's understanding of information that was not explicitly stated in the picture description they heard. For example, they describe the explicit recall and implicit comprehension evaluated in one of their comics, saying: "the placement of a bench was explicitly mentioned in relation to where a dog is in the image. . . the dog's placement is explicitly described in relation to a woman whereas the location of the woman in relation to the bench is logically implied via her relationship to the dog." (p. 820). In order to go beyond investigating explicit recall of specific items that were mentioned in the narratives of the current study, it was decided to also assess their discourse comprehension in terms of the semantic relationship between the different elements (i.e., the narrative's event structure, see Li et al., 2017). This is distinguished from recall in that while recall tests participants' ability to retrieve lexical information regarding elements in the story (e.g., the presence of a cat, a television, and 3 birds in the narrative), discourse comprehension measures participant's understanding of the relationship between these items (e.g., that the cat is **trying to catch** the three birds **that are being televised on the screen**, which ultimately leads to the cat **breaking** the television). As such, each drawing was evaluated on a Likert scale for the general comprehension of the event structure of the narrative. The Likert scale was on a scale of zero to five, where 0 corresponded to absolutely no correspondence between the drawings and the narrative, to 5 indicating a complete understanding of the event structure of the story (see **Table 3**). See section "**Appendix C**" for an example scoring of discourse comprehension. Thus each drawing was given a recall score for each gesturally marked element in the narrative, and one single score for discourse comprehension.

Reliability

Interrater reliability was calculated using Fleiss' kappa with three additional raters evaluating both recall and comprehension for a total of 64 drawings, representing 18.6% of all the data. The calculation of recall scores were based on evaluators' individual

TABLE 3 | The scoring rubric to evaluate comprehension.

Score	Interpretation	Description
0	Not-evaluable	The drawing had no correspondence with any aspect of the narrative or was left blank
1	No understanding of the narrative	Perhaps drew a character or object, but no story development is present
2	Minimal understanding of the narrative	Drew at least one event from the narrative, but minimal story development
3	Partial understanding of the narrative	Drew multiple events from the narrative, understands at least partially the "main goal" but misunderstands some other aspects of the narrative
4	Near complete understanding of the narrative	Clearly understood main goal of the narrative, as well as possibly some other minor details that are implicated in the story
5	Complete understanding of the narrative	Clearly understood the main goal of the narrative, as well as other minor details that are implicated in the story

scores for each gesturally marked item (where a score of 2 indicates perfect recall, a score of 1 indicates distorted recall or ambiguity, and a score of 0 indicates no recall, see section "Scoring of the drawing task"). Fleiss' kappa showed that there was good agreement between the raters' scores, κ = 0.713 (95% CI, 0.713 to 0.714, $p < 0.001$).

In terms of comprehension, reliability was calculated using the individual comprehension scores. Fleiss' kappa showed moderate agreement between the raters, κ = 0.529 (95% CI, 0.527 to 0.531, $p < 0.001$). Reliability was further calculated by grouping the individual comprehension scores so that a score of 1 or 2 would be binned as "low comprehension" and a score of 4 or 5 would be binned as "high comprehension." Fleiss' kappa showed good agreement between the raters, κ = 0.723 (95% CI, 0.720 to 0.725, $p < 0.001$).

Statistical Analyses

Two Generalized Linear Mixed Models (GLMMs) were applied to the recall and comprehension scores using the *glmmTMB* package in R (Brooks et al., 2017). For both GLMMs, the fixed factors were Condition (two levels: Gesture and No Gesture), Language (two levels: Native and Non-native) as well as their interaction. To determine the random effects structure for each GLMM, a series of Linear Mixed Models were modeled using all the potential combinations of random effects, from the most complex structure to a basic model containing no random effects. Structures that did not produce any converge problems were then compared using the "compare performance" function from the *performance* package (Lüdecke and Makowski, 2019) to identify the best fitting model for the data. In other words, this process assesses all of the possible random effects structures and returns the best-fitting model. For both dependent variables, the best fitting model was a random effects structure which included a random intercept for item (i.e., the individual comic narrative) and a random slope for Language by Participant. Omnibus test results are described below, as well as the results from a series of Bonferroni pairwise tests carried out with the *emmeans* package (Lenth, 2019), which includes a measure of effect size (via Cohen's d).

RESULTS

Figure 3 below shows the average recall score (in%) for both Language and Gesture Conditions. Results of the GLMM with recall score as the dependent variable reveal a significant main effects of Language [$\chi 2(1) = 88.297$, $p < 0.001$] and Condition [$\chi 2(1) = 5.248$, $p = 0.022$], as well as a significant interaction between Language and Condition [$\chi 2(1) = 4.150$, $p = 0.042$]. *Post hoc* comparisons showed that participants did significantly better in Native listening conditions than in Non-native listening conditions ($d = -1.83$, $p < 0.001$) and did significantly better in the No-Gesture condition than the Gesture condition ($d = -0.25$, $p = 0.023$). As for the significant interaction, while gesture had no impact on recall in Native listening conditions ($d = -0.03$, $p = 0.855$), participants scored significantly better in the No-Gesture condition than in the Gesture condition when in Non-native listening conditions ($d = -0.47$, $p = 0.002$). From these results, it seems that while beat gesture has no major effect for native listeners, they negatively impact recall when participants listen to a non-native language.

Figure 4 below shows the average comprehension score for both Language and Gesture Conditions. Results of the GLMM with comprehension score as the dependent variable reveal a significant main effects of Language [$\chi 2(1) = 68.398$, $p < 0.001$] and a significant interaction between Language and Condition [$\chi 2(1) = 9.673$, $p = 0.002$]. Similar to the recall scores, *post hoc* comparisons showed that participants did significantly better in Native listening conditions than in Non-native listening conditions [$d = -1.84$, $p < 0.001$]. In regards to the interaction, while gesture had no impact on comprehension in Native listening conditions [$d = 0.18$, $p = 0.249$], participants scored significantly better in the No-Gesture condition than in the Gesture condition when in Non-native listening conditions [$d = -0.49$, $p = 0.001$]. Thus similar to the results on recall, it seems that while beat gesture has no major effect on comprehension for native listeners, they negatively impact comprehension when participants listen to a non-native language.

When comparing the recall and comprehension scores regardless of condition, we find a significant, positive correlation between the two scores [$r(342) = 0.893$, $p < 0.001$], suggesting that as participants remembered more individual items in the narratives, they also better understood the overall event structure of the narrative.

DISCUSSION

The results of the present investigation show that while the presence or absence of beat gestures in discourse does not affect either recall or comprehension of complex narrative speech for

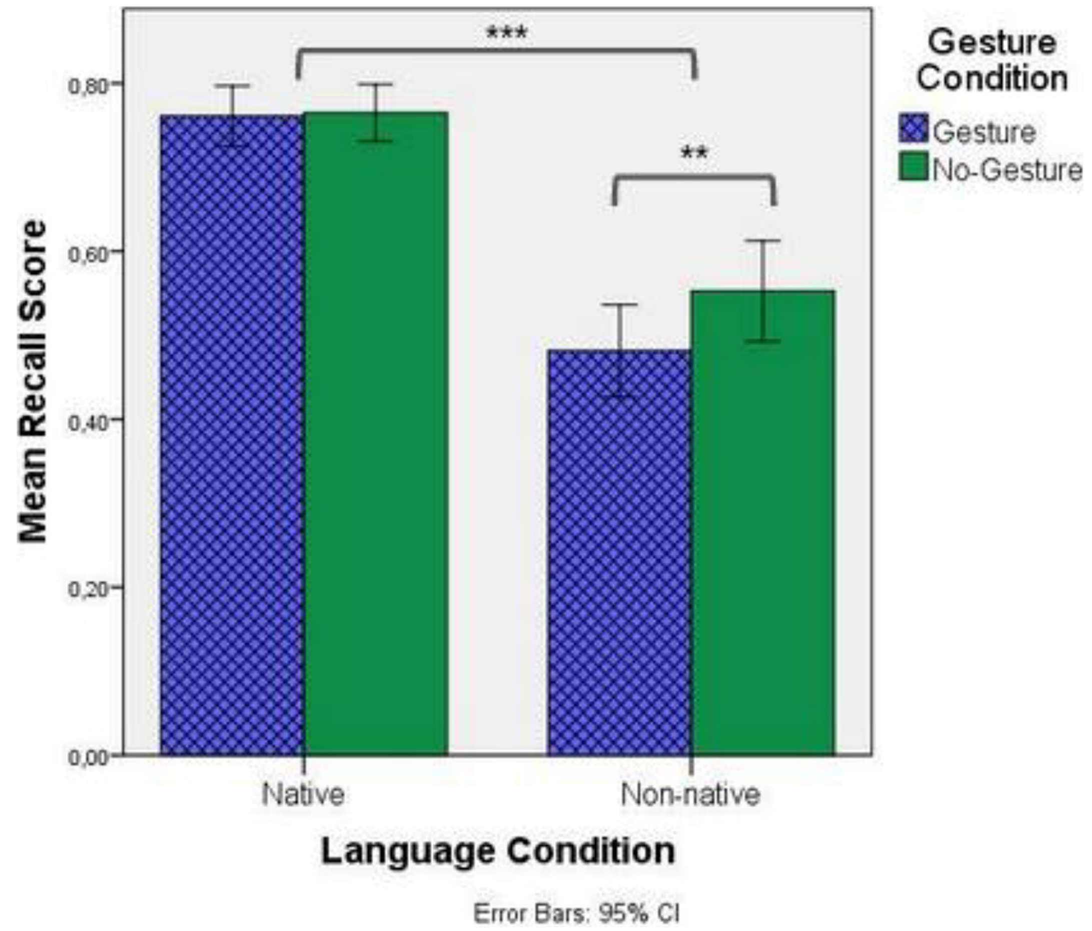


FIGURE 3 | Mean recall scores by Language and Gesture conditions. "**" Refers to a *p*-value less than 0.01, while "***" refers to a *p*-value less than 0.001.

native listeners, when those same listeners are exposed to speech that is not in their native language and of which they have an intermediate proficiency level, non-referential beat gestures significantly impede both recall and comprehension.

First, the results in terms of the non-beneficial effects of non-referential beat gesture on native language contexts contribute to expand and refine our knowledge about the benefits of gesture in recall and comprehension processes and further understand some of the reasons behind the conflicting results. Our results are in line with results from the studies by Dahl and Ludvigsen (2014) and Austin and Sweller (2014), where neither study found any benefit of gestures (referential gestures in the case of the former, neither referential nor non-referential in the case of the latter) for information recall. Importantly these results contrast with other studies that report positive results for both of these gestures. By looking closely at the stimuli of the two studies it is particularly interesting to note that methodologically these reflect the methodology in the current study in terms of the stimuli used. Particularly regarding the length of the narratives and the number of non-referential gestures used, the current study as well as both Austin and Sweller's (2014) and Dahl and Ludvigsen's (2014) studies were similar. Interestingly, the stimuli were substantially longer and contained more gestures than studies that found positive effects (e.g., Kushch and Prieto, 2016). Thus a potential reason that these gestures do not boost recall and comprehension is gesture rate, i.e., the fact that speakers repeatedly used gestures (in our study, between two to three lexical items were marked with a gesture per sentence, see **Table 1**).

Thus our interpretation of the non-beneficial effects of non-referential beat gestures in the native speaker group is that having a high rate of gesture may have "bleached" their pragmatic intent, provoking changes in the listener's processing of discourse. By contrast, previous evidence has shown that when non-referential beat gestures occur with the specific pragmatic function of contrastive focus (e.g., Wang and Chu, 2013; Dimitrova et al., 2016; Kushch and Prieto, 2016; Llanes-Coromina et al., 2018; Morett and Fraundorf, 2019) or highlighting one of the items in a list (Igualada et al., 2017), these gestures are beneficial for recall or comprehension. In the current study, the speakers after which the target stimuli were modeled were instructed to "speak expressively with their hands" which may have ultimately led to an exaggerated performance in terms of the number of non-referential beat gestures that were produced. This increase in the number of gestures may have hidden any real

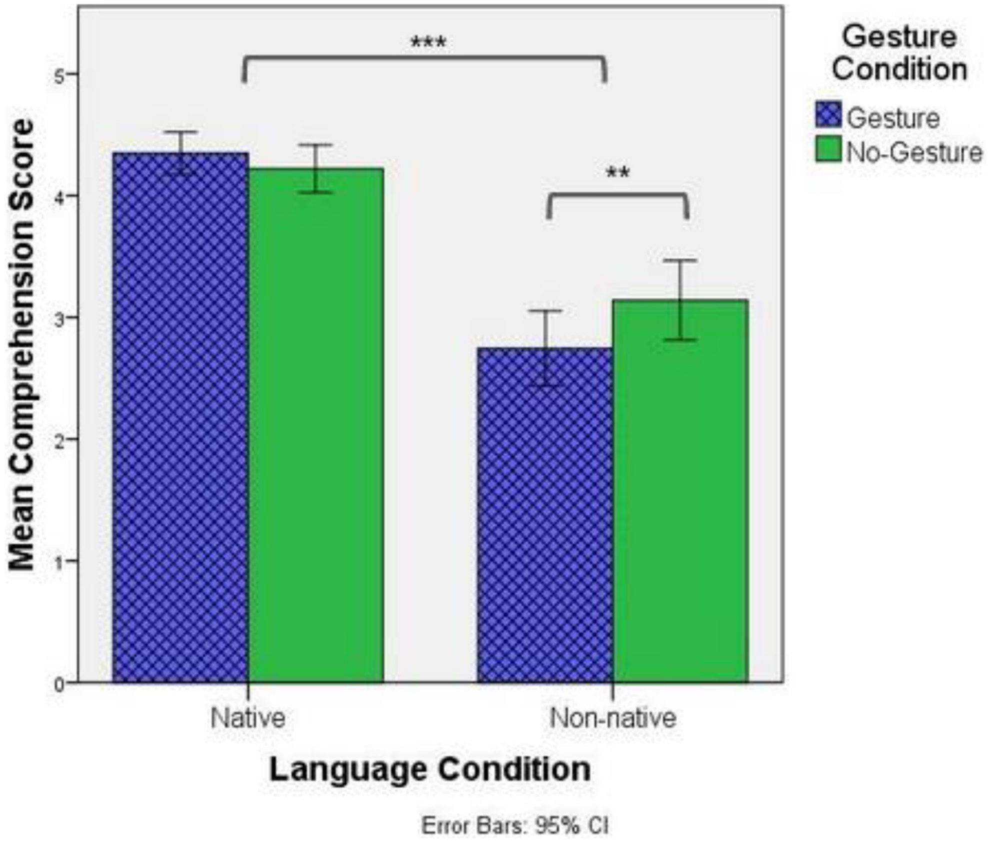


FIGURE 4 | Mean comprehension scores by Language and Gesture condition. "**" Refers to a *p*-value less than 0.01, while "***" refers to a *p*-value less than 0.001.

pragmatic relevance to them, ultimately using non-referential beats that were no longer pragmatically relevant. Most of the non-referential beat gestures that were produced in our target narrations marked information structure (i.e., new referents, broad focus, narrow focus, etc.). That is, they marked information that the speaker would have deemed "important." However, it might well be that in marking too many elements as important in discourse, the inherent property of marking something as separate (i.e., "important") is reduced, ultimately reducing the effectiveness of non-referential beat gestures as highlighters of important information (McNeill, 1992; see also Biau and Soto-Faraco, 2013; Dimitrova et al., 2016, among others). This is also in direct contrast with studies that showed benefits in semantic integration and comprehension (e.g., Wang and Chu, 2013; Llanes-Coromina et al., 2018), where the presence of a beat gesture on a contrastively marked element may have increased the listener's interpretation of speaker certainty, reducing doubt in their interpretation of speech and ultimately aiding in semantic processing. As the current study did not use gestures to merely mark contrastive elements, they may not have had this effect of reducing the certainty of the listener's semantic interpretation.

Parallels of what we can classify as a *gesture rate effect* can be drawn from the interpretation of typographic prominence (e.g., capital letters). Scott and Jackson (2020) describe how using capitalized letters in the written modality can give the reader an impression of emphasis. However, a stylistic choice to write entirely in capital letters causes the reader to no longer interpret capitalization as a marker of emphasis and thus must do something different to mark emphasis (e.g., putting an emphasized element in italics). Thus, it is sensible to conclude that whenever beat gestures are repeatedly used in discourse, they inherently lose their saliency as markers of important information.

Moreover, presumably the fact that repeatedly used beat gestures triggered not only a loss of their pragmatic saliency but also potentially led our listeners to adapt their reliance on gesture based on speaker style. Indeed, two studies have already shown how listeners adapt to the gestural behavior of their interlocutor. The previously mentioned EEG study by Obermeier et al. (2015) showed that when listeners see speakers producing both meaningless grooming gestures along with iconic gestures, they do not process their iconic gestures as strongly as when speakers did not perform any grooming gestures. Similarly, a recent behavioral study with beat gestures by Morett and Fraundorf (2019) defended a top-down approach in discourse processing. This "top-down" approach implies that listeners attune to the gestural habits of speakers and make inferences about their intentions based on their behavior (as opposed to a bottom-up approach where merely the presence of cues in the speech signal guide the listener's interpretation). Within the interpretation of these studies, it seems as though the native-listeners were exposed to repeatedly produced beat gestures, making these gestures unreliable and ultimately failing to raise attention to important information in speech and reducing any potential benefit for recall.

Second, along with recent studies on referential gestures, the results of the present investigation showed that beats had negative effects for low-intermediate language learners. Our results complement and expand previous findings showing that lower-level language learners show increased processing cost when gestures are present and that gesture processing stops when speech becomes too difficult to understand (Ibáñez et al., 2010; Drijvers and Özyürek, 2018). In terms of our results, participants may have been at a disadvantage from increased processing costs for gesture, doubled with the lack of semantic information to be gleaned from these movements. As such, perhaps the non-native listeners at a low-intermediate level are still dependent on clear semantic meaning in gestures. By contrast, the studies by Dahl and Ludvigsen (2014) and Drijvers et al. (2019a), who found positive effects of iconic gestures on recall and comprehension processes, recruited advanced learners and exposed them to referential gestures, whereas in the current study, the non-native listeners had a low-intermediate level and were exposed to non-referential gestures.

Our study is not the first to find negative effects for gestures. In terms of L2 novel word learning, Kelly and Lee (2012) found that when teaching word pairs that differ by only a geminate, the presence of referential gestures had a negative effect on the participants' word learning. However, the gestures were indeed beneficial whenever the word pair differed by both a geminate and their segmental composition. The authors thus suggest that gestures are only helpful when phonetic demands are low. Another study using an electrophysiological paradigm by Zhang et al. (2020) used naturalistic stimuli to investigate how multimodal cues interact in discourse processing, notably the N400. This study particularly stands out, as they used natural stimuli that contained multiple gestures (and often beat gestures). Interestingly, they found that when controlling for linguistic surprisal for each word, referential gestures had a tendency to lower the N400 (generally interpreted as easing semantic integration), while beat gestures tended to have the opposite effect.

The findings from the current study are limited in a few aspects. First, the actresses that were featured in the stimuli were given no specific instructions in terms of prosody in order to maintain the naturalness of the stimuli (i.e., to avoid having to overlay audio tracks and blur faces, etc.). While beat gestures tend to associate with speech prominence, studies have shown that the production of a beat gesture affects how acoustic prominence is realized in speech (e.g., Krahmer and Swerts, 2007; Pouw et al., 2020). Thus it is possible that differences in the phonetic realization of prominence may have had an effect. Conversely, other studies have also shown that when prosody is held constant, the presence of a beat gesture boosts the perception of speech prominence (Krahmer and Swerts, 2007; Bosker and Peeters, 2020). Even though our materials were controlled for the presence of pitch accentuation in beat positions across conditions, the fact that speech production was not kept completely constant does not rule out the possibility that pitch range differences might have had an effect on the results. Thus, future studies should control for phonetic differences in prosodic prominences to flush out to what extent it is modulation in the visual or auditory cues to prominence that are the driving factor behind these effects.

Another limitation of this study regards the methodological choices. The study only looked at intermediate learners of English. By adding high proficient learners, it would have been possible to flush out any proficiency-level effects. This could potentially show at what stage in learning non-referential gestures stop being detrimental for recall and comprehension in language learners. Another limitation is in regards to the processing costs of our participants. Also, by adding an electrophysiological element to the study, we would have been able to directly measure these processing costs. The task itself may have been a limiting factor, particularly for participants who did not feel confident in drawing. Though participants were reassured by the experimenter that their drawings could be simple stick figures and that they could write words and draw arrows for things that may have been difficult to draw, and all of them expressed enough confidence in an informal way, it would be good for future studies to take a measure of drawing confidence in the task and factor this variable into the statistical modeling.

Finally, it is important also to consider that all of the participants in the current study were native French learners of English. As such, we cannot discard the possibility of L1 language effects in the results of the effects of beats in the L2. While non-referential beat gestures show similar patterns of integration with speech prominence in both languages (see Shattuck-Hufnagel and Ren, 2018; Rohrer et al., 2019), in terms of focus marking, French makes use of thematic structures more often than prosodic focus, and non-referential beat gestures tend to align more with the prosodic focusing than thematic structures (Ferré, 2011). In other words, native French listeners may rely less on prosodic and gestural marking for focus. English, on the other hand does not use clefting strategies as often to mark focus (e.g., Vander Klok et al., 2018), potentially making beat gestures a more reliable marker of focus than in French. As such, it would be interesting to see if similar results were found with native English learners of French, or in a completely different pair of languages. In the case that there is no difference between populations, inherent language differences could be ruled out.

All in all, the current study adds to our understanding of the role of gesture in recall and comprehension processes by giving insight into *when* gestures are beneficial for listeners, both native and non-native. Methodologically, the results of our study highlight the need for researchers to take task complexity into account when interpreting results on gesture-speech integration processes, and particularly the effects the length of the discourse, the pragmatic functions of gesture, and the gesture rate. This is particularly true in second language contexts. While previous positive results could have led language instructors to believe that adding non-referential beat gestures to their discourse would be beneficial for their students, results from the current study suggest that this is not necessarily the case and that degree of proficiency and task complexities are important factors that need to be taken into account. Instructors are encouraged to reflect more on using beat gestures in specific, relevant contexts and to select precisely what information is important for the listener, and finally take into account that level of proficiency in the foreign language is a crucial factor in the processing of gesture-speech integration.

DATA AVAILABILITY STATEMENT

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

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. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

PR, ED-R, and PP contributed equally to the development of the research questions, the experimental design, and the discussion of the results. PR carried out the data collection and analysis, and was in charge of the writing of the article, with feedback from ED-R and PP. All authors contributed to the article and approved the submitted version.

FUNDING

The first author would like to acknowledge a joint PhD grant, awarded by the Department of Translation and Language Sciences, Universitat Pompeu Fabra, and SGR AGAUR Grant, Generalitat de Catalunya, Award number: 2017 SGR_971. Spanish Ministry of Science, Innovation and Universities, Grant/Award Number: PGC2018-097007-B-I00. SGR AGAUR Grant, Generalitat de Catalunya, Award number: 2017 SGR_971. The GEHM (Gesture and Head Movements in Language) Research Network, funded by the Independent Research Fund Denmark, Grant/Award number: 9055-00004B.

ACKNOWLEDGMENTS

The authors would like to thank the students and staff at the Université de Nantes who made this research possible (particularly Emmanuella Annan, Tatiana Soldatova, Anne-Yaëlle Fleurant, and David Gregory). A special thanks goes to those who graciously volunteered their time for the creation and evaluation of the experimental stimuli, and the actresses who performed them. A special thanks also goes to Joan Borràs-Comes for help with the statistical analyses, as well as the feedback from colleagues both at the LLING lab in Nantes, and the GrEP research group in Barcelona.

REFERENCES

Austin, E. E., and Sweller, N. (2014). Presentation and production: the role of gesture in spatial communication. *J. Exp. Child Psychol.* 122, 92–103. doi: 10.1016/j.jecp.2013.12.008

Biau, E., and Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain Lang.* 124, 143–152. doi: 10.1016/j.bandl.2012.10.008

Bosker, H. R., and Peeters, D. (2020). Beat gestures influence which speech sounds you hear. *bioRxiv* [Preprint]. doi: 10.1101/2020.07.13.200543

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., et al. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* 9, 378–400.

Cohen, R. L., and Otterbein, N. (1992). The mnemonic effect of speech gestures: pantomimic and non-pantomimic gestures compared. *Eur. J. Cogn. Psychol.* 4, 113–139. doi: 10.1080/09541449208406246

Cravotta, A., Busà, M. G., and Prieto, P. (2019). Effects of encouraging the use of gestures on speech. *J. Speech Lang. Hear. Res.* 62, 3204–3219. doi: 10.1044/2019_JSLHR-S-18-0493

Dahl, T. I., and Ludvigsen, S. (2014). How I see what you're saying: the role of gestures in native and foreign language listening comprehension. *Mod. Lang. J.* 98, 813–833. doi: 10.1111/modl.12124

Dargue, N., and Sweller, N. (2020). Two hands and a tale: when gestures benefit adult narrative comprehension. *Learn. Instr.* 68:101331. doi: 10.1016/j.learninstruc.2020.101331

Dimitrova, D., Chu, M., Wang, L., Özyürek, A., and Hagoort, P. (2016). Beat that word: how listeners integrate beat gesture and focus in multimodal speech discourse. *J. Cogn. Neurosci.* 28, 1255–1269. doi: 10.1162/jocn_a_00963

Drijvers, L., and Özyürek, A. (2017). Visual context enhanced: the joint contribution of iconic gestures and visible speech to degraded speech comprehension. *J. Speech Lang. Hear. Res.* 60, 212–222. doi: 10.1044/2016_JSLHR-H-16-0101

Drijvers, L., and Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain Lang.* 177–178, 7–17. doi: 10.1016/j.bandl.2018.01.003

Drijvers, L., Vaitonytë, J., and Özyürek, A. (2019a). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cogn. Sci.* 43:e12789. doi: 10.1111/cogs.12789

Drijvers, L., van der Plas, M., Özyürek, A., and Jensen, O. (2019b). Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. *Neuroimage* 194, 55–67. doi: 10.1016/j.neuroimage.2019.03.032

Ferré, G. (2011). "Thematisation and prosodic emphasis in spoken French," in *Proceedings of the Gestures and Speech in Interaction, GESPIN*, Bielefeld, 1–6.

Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: their meaning matters. *Eur. J. Cogn. Psychol.* 18, 185–205. doi: 10.1080/09541440540000158

Holle, H., and Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *J. Cogn. Neurosci* 19, 1175–1192. doi: 10.1162/jocn.2007.19.7.1175

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol. Bull.* 137, 297–315. doi: 10.1037/a0022128

Ibáñez, A., Manes, F., Escobar, J., Trujillo, N., Andreucci, P., and Hurtado, E. (2010). Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neurosci. Lett.* 471, 48–52. doi: 10.1016/j.neulet.2010.01.009

Igualada, A., Esteve-Gibert, N., and Prieto, P. (2017). Beat gestures improve word recall in 3-to 5-year-old children. *J. Exp. Child Psychol.* 156, 99–112. doi: 10.1016/j.jecp.2016.11.017

Im, S., and Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proc. Linguist. Soc. Am.* 5:685. doi: 10.3765/plsa.v5i1.4755

Kelly, S. D., Creigh, P., and Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254

Kelly, S. D., and Lee, A. L. (2012). When actions speak too much louder than words: hand gestures disrupt word learning when phonetic demands are high. *Lang. Cogn. Process.* 27, 793–807. doi: 10.1080/01690965.2011.581125

Kelly, S. D., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Lang. Cogn. Process.* 24, 313–334. doi: 10.1080/01690960802365567

Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005

Kushch, O., Igualada, A., and Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Lang. Cogn. Neurosci.* 33, 992–1004. doi: 10.1080/23273798.2018.1435894

Kushch, O., and Prieto, P. (2016). "The effects of pitch accentuation and beat gestures on information recall in contrastive discourse," in *Proceedings of the Speech Prosody 2016, Boston, MA*, eds J. Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux (Vientiane: International Speech Communication Association), 922–925.

Lenth, R. (2019). *Estimated Marginal Means, Aka Least-Squares Means. R package Version 1.3.4.*

Levantinou, E. I., and Navarretta, C. (2015). "An investigation of the effect of beat and iconic gestures on memory recall in L2 speakers," in *Proceedings of the 3rd European Symposium on Multimodal Communication, Dublin*, eds E. Gilmartin, L. Cerrato, and N. Campbell (Linköping: Linköping University Electronic Press), 32–37.

Li, X., Zhang, Y., Zhao, H., and Du, X. (2017). Attention is shaped by semantic level of event-structure during speech comprehension: an electroencephalogram study. *Cogn. Neurodyn.* 11, 467–481. doi: 10.1007/s11571-017-9442-4

Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borràs-Comes, J., and Prieto, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. *J. Exp. Child Psychol.* 172, 168–188. doi: 10.1016/j.jecp.2018.02.004

Lüdecke, D., and Makowski, D. (2019). *Performance: Assessment of Regression models Performance. R Package Version 0.1.0.*

Macedonia, M., Müller, K., and Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Hum. Brain Mapp.* 32, 982–998. doi: 10.1002/hbm.21084

Macoun, A., and Sweller, N. (2016). Listening and watching: the effects of observing gesture on preschoolers' narrative comprehension. *Cogn. Dev.* 40, 68–81. doi: 10.1016/j.cogdev.2016.08.005

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago press.

Morett, L. M. (2014). When hands speak louder than words: the role of gesture in the communication, encoding, and recall of words in a novel second language. *Mod. Lang. J.* 98, 834–853. doi: 10.1111/modl.12125

Morett, L. M., and Fraundorf, S. H. (2019). Listeners consider alternative speaker productions in discourse comprehension and memory: evidence from beat gesture and pitch accenting. *Mem. Cogn.* 47, 1515–1530. doi: 10.3758/s13421-019-00945-1

Obermeier, C., Holle, H., and Gunter, T. C. (2011). What iconic gesture fragments reveal about gesture-speech integration: when synchrony is lost, memory can help. *J. Cogn. Neurosci.* 23, 1648–1663. doi: 10.1162/jocn.2010.21498

Obermeier, C., Kelly, S. D., and Gunter, T. C. (2015). A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration. *Soc. Cogn. Affect. Neurosci.* 10, 1236–1243. doi: 10.1093/scan/nsv011

Pouw, W., Harrison, S. J., and Dixon, J. A. (2020). Gesture-speech physics: the biomechanical basis for the emergence of gesture-speech synchrony. *J. Exp. Psychol. Gen.* 149, 391–404. doi: 10.1037/xge0000646

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P. L., and Vilà-Giménez, I. (2018). "Deconstructing beat gestures: a labelling proposal," in *Proceedings of the 9th International Conference on Speech Prosody 2018, Poznañ*, eds K. Klessa, J. Bachan, A. Wagner, M. Karpiñski, and D. Śledziñski (Vientiane: International Speech Communication Association), 201–205. doi: 10.21437/SpeechProsody.2018-41

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: three experiments exploring a neglected facet of communication. *J. Nonverbal Behav.* 5, 172–183. doi: 10.1007/BF00986134

Rohrer, P. L., Prieto, P., and Delais-Roussarie, E. (2019). "Beat gestures and prosodic domain marking in French," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, VIC*, eds S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Canberra, ACT: Australasian Speech Science and Technology Association Inc), 1500–1504.

Scott, K., and Jackson, R. (2020). "When everything stands out, nothing does," in *Relevance Theory, Figuration, and Continuity in Pragmatics*, ed. A. Piskorska (Amsterdam: John Benjamins), 167–192. doi: 10.1075/ftl.8.06sco

Shattuck-Hufnagel, S., and Prieto, P. (2019). "Dimensionalizing co-speech gestures," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, VIC*, eds S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Canberra, ACT: Australasian Speech Science and Technology Association Inc), 1490–1494.

Shattuck-Hufnagel, S., and Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Front. Psychol.* 9:1514. doi: 10.3389/fpsyg.2018.01514

So, W. C., Sim Chen-Hui, C., and Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: is meaning in gesture important for memory recall? *Lang. Cogn. Process.* 27, 665–681. doi: 10.1080/01690965.2011.573220

Sueyoshi, A., and Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Lang. Learn.* 55, 661–699. doi: 10.1111/j.0023-8333.2005.00320.x

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture* 8, 219–235. doi: 10.1075/gest.8.2.06tel

Vander Klok, J., Goad, H., and Wagner, M. (2018). Prosodic focus in English vs. French: a scope account. *Glossa J. Gen. Linguist.* 3:71. doi: 10.5334/gjgl.172

Wang, L., and Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia* 51, 2847–2855. doi: 10.1016/j.neuropsychologia.2013.09.027

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., and Vigliocco, G. (2020). More than words: the online orchestration of word predictability, prosody, gesture, and mouth movements during natural language comprehension. *bioRxiv* [Preprint]. doi: 10.1101/2020.01.08.896712

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.

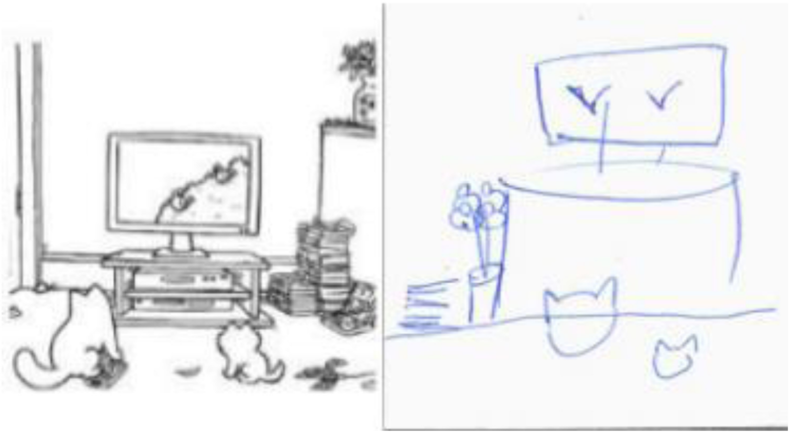
APPENDIX A: NARRATION FOR THE COMIC IN FIGURE 1 IN ENGLISH AND FRENCH (BOLD, CAPITALIZED WORDS INDICATE A GESTURALLY MARKED ELEMENT)

FIRST, we see a **LARGE** television on a **TABLE**. On the **TELEVISION** screen, we see **TWO** birds. To the **RIGHT** of the television there are two **STACKS** of magazines on the **FLOOR**, and a **SHELF** with a **FLOWER** vase. A **CAT** and a **SMALL** kitten are sitting in **FRONT** of the **TV**, watching the two birds. The cat, which is on the **LEFT**, has a remote control under his **RIGHT** hand. **NEXT**, we see that the television screen shows **THREE** birds, and both the **CAT** and the **KITTEN** have climbed **ONTO** the table. The **LARGER** cat, now on the **RIGHT**, is reaching his hand toward the screen. **THEN**, we see the **TV** screen shows **ONE** bird on a **BRANCH**. The large cat has climbed on **TOP** of the **TV** screen and is looking **DOWN** at the bird, while the kitten is **UNDER** the large cat, **BEHIND** the television. **FINALLY**, we see the **TWO** cats on the table, looking **SHOCKED**. The **TV** has **FALLEN** to the ground and the back of the television is **CRACKED**.

D'ABORD, on voit un **GRAND** téléviseur sur un **MEUBLE**. Sur **L'ECRAN** de la télé, on voit **DEUX** oiseaux. A **DROITE**, il y a **DEUX** piles de magazines et une **ÉTAGÈRE** avec un vase de **FLEURS** dessus. Il y a un **GROS** chat et un **PETIT** chat assis **DEVANT** la télé, en train de regarder les deux oiseaux. Le gros chat, à **GAUCHE**, tient la télécommande dans sa main **DROITE**. **ENSUITE**, on voit sur l'écran **TROIS** oiseaux, et les **DEUX** chats sont montés sur le **MEUBLE**. Le **GROS** chat, maintenant à **DROITE**, lève la main vers l'écran. **PUIS**, on voit sur l'écran **UN** oiseau sur une **BRANCHE**. Le gros chat est **MONTÉ** sur la télé et regarde en **BAS** vers l'oiseau, alors que le petit chat est **DERRIÈRE** la télé. **FINALEMENT**, on voit les **DEUX** chats sur le meuble, l'air **CHOQUÉ**. La télé est **TOMBÉE** par terre et l'arrière de la télé est **CASSÉ**.

APPENDIX B: EXAMPLE OF RECALL EVALUATION

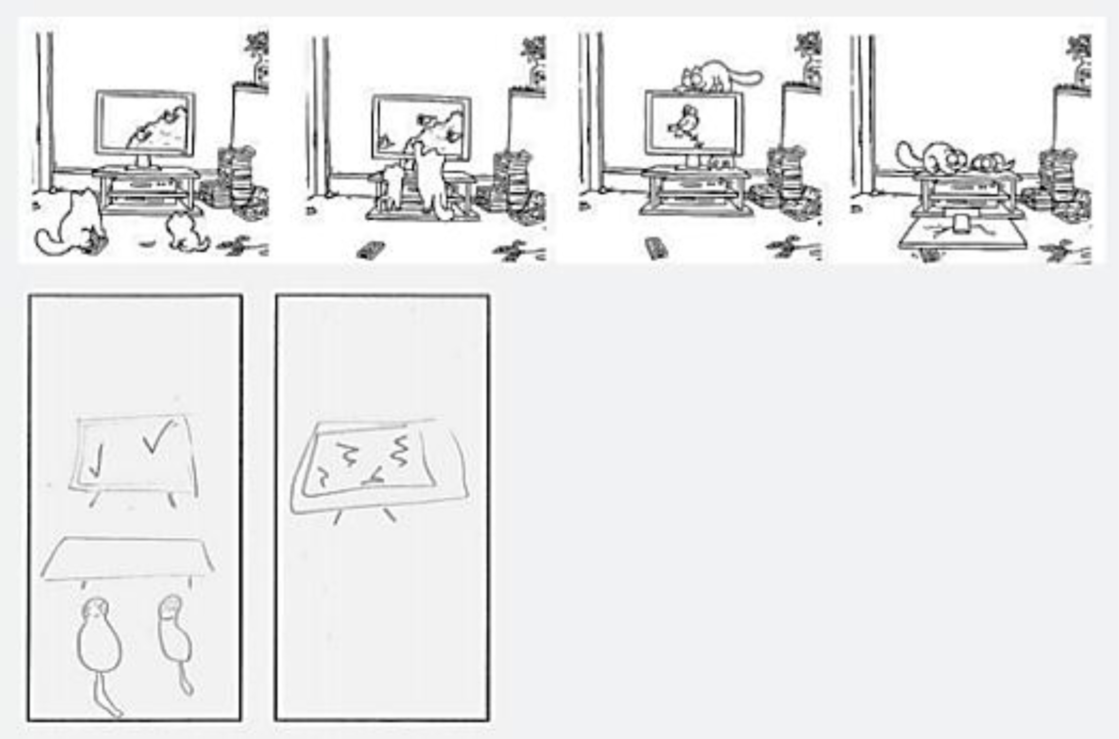
The left panel of the image below shows the original comic illustration, and the panel on the right shows the illustration provided by the participant. The table shows the number of points given for each gesturally marked element (**bold** indicates words that are gesturally marked elements, underline indicates the gesturally marked element being evaluated).



Element	Points
There are **two** birds on the **television** screen	2
There are **two** birds on the **television** screen	2
There are two **stacks** of magazines on the **floor**	1
There are two **stacks** of magazines on the **floor**	2
There is a **shelf** with a flower **vase**	0
There is a **shelf** with a flower **vase**	2

APPENDIX C: EXAMPLE OF DISCOURSE COMPREHENSION EVALUATION

The upper panel of the image below shows the original comic illustration, and the lower panel shows the illustration provided by the participant. The participant's illustration demonstrates that in terms of recall, specific gesturally marked items were remembered, however, the general understanding of the narrative's event structure is lacking. The participant drew one action (the TV breaking), yet did not include information regarding what caused the TV to break (i.e., the cats climbing on the TV, trying to catch the birds on the screen). This suggests that the participant did not understand how the cats were implicated in the narrative. The participant drew a few objects related to the story, and one action (the TV breaking), so this participant received a score of 2 for discourse comprehension.



frontiers in Psychology

BRIEF RESEARCH REPORT
published: 22 October 2020
doi: 10.3389/fpsyg.2020.575991

Semantic Relationships Between Representational Gestures and Their Lexical Affiliates Are Evaluated Similarly for Speech and Text

Sarah S. Hughes-Berheim†*, Laura M. Morett**†* and Raymond Bulger*

Department of Educational Studies in Psychology, Research Methodology and Counseling, University of Alabama, Tuscaloosa, AL, United States

This research examined whether the semantic relationships between representational gestures and their lexical affiliates are evaluated similarly when lexical affiliates are conveyed via speech and text. In two studies, adult native English speakers rated the similarity of the meanings of representational gesture-word pairs presented via speech and text. Gesture-word pairs in each modality consisted of gestures and words matching in meaning (semantically-congruent pairs) as well as gestures and words mismatching in meaning (semantically-incongruent pairs). The results revealed that ratings differed by semantic congruency but not language modality. These findings provide the first evidence that semantic relationships between representational gestures and their lexical affiliates are evaluated similarly regardless of language modality. Moreover, this research provides an open normed database of semantically-congruent and semantically-incongruent gesture-word pairs in both text and speech that will be useful for future research investigating gesture-language integration.

Keywords: representational gesture, gesture comprehension, gesture-text relationship, gesture-speech relationship, Integrated Systems Hypothesis

OPEN ACCESS

Edited by:
Autumn Hostetter,
Kalamazoo College, United States

Reviewed by:
Elizabeth Wakefield,
Loyola University Chicago,
United States
Elizabeth Austin,
Macquarie University, Australia

****Correspondence:***
Laura M. Morett
lmorett@ua.edu
orcid.org/0000-0002-1251-7213

† These authors have contributed equally to this work

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

Received: *28 June 2020*
Accepted: *25 September 2020*
Published: *22 October 2020*

Citation:
Hughes-Berheim SS, Morett LM and Bulger R (2020) Semantic Relationships Between Representational Gestures and Their Lexical Affiliates Are Evaluated Similarly for Speech and Text. Front. Psychol. 11:575991.
doi: 10.3389/fpsyg.2020.575991

INTRODUCTION

Gesture can be defined as hand or body movements that convey information (Özyürek, 2002; Melinger and Levelt, 2005). Most gestures are gesticulations (hereafter referred to simply as "gestures"), which are naturally produced in conjunction with speech (see Hostetter, 2011, for a review). According to McNeill, 1992, 2005 gesture taxonomy, deictic gestures indicate presence (or absence) of objects via pointing; beat gestures convey speech prosody and emphasis; and representational (i.e., metaphoric and iconic) gestures convey meaning relevant to co-occurring speech via form and motion. Representational gestures may be used to describe actions (e.g., swinging a bat), to depict spatial properties (e.g., describing a ring as round), or to refer to concrete entities associated with abstract ideas (e.g., putting a hand over one's heart to convey love; Hostetter, 2011). Gesturing while speaking is so pervasive that gesture and speech have been argued to be inextricably integrated into mental representations of language (Kendon, 2000). The process of producing speech and gesture is thought to occur bi-directionally, such that speech production influences gesture production, and conversely, gesture production influences speech production (Kita and Özyürek, 2003).

By the same logic, gesture and speech are similarly integrated during language comprehension. The Integrated Systems Hypothesis (Kelly et al., 2010) posits that co-occurring gesture and speech interact bi-directionally during language processing to enhance comprehension. This interaction occurs obligatorily, such that information from one modality (speech) cannot be processed without being influenced by information from the other modality (gesture). This hypothesis is supported by behavioral findings indicating fast and accurate identification of an action in a prime video followed by a target video displaying semantically-congruent representational gesture and speech related to the prime. In contrast, identification of action in a prime video is relatively slow and inaccurate when it is followed by a target video containing gesture, speech, or both that are semantically-incongruent and partially unrelated to the prime. Further, even if instructions are issued to attend to speech and ignore accompanying gesture, error rates are higher when prime and target videos are semantically-incongruent than when they are semantically-congruent (Kelly et al., 2010).

The bi-directional and obligatory integration of gesture and speech postulated by the Integrated Systems Hypothesis has important implications for learning. Comprehension accuracy and speed are bolstered by viewing semantically-congruent representational gestures accompanying speech (Drijvers and Özyürek, 2017, 2020). Moreover, words learned with semantically-congruent representational gestures are remembered more accurately than words learned without gestures (Kelly et al., 2009; So et al., 2012). In addition to supporting the Integrated Systems Hypothesis, these findings are consistent with Dual Coding Theory (Clark and Paivio, 1991), which posits that representational gesture splits the cognitive load between the visual and verbal representational systems, freeing up cognitive resources and thereby enhancing comprehension. These findings suggest that when novel vocabulary is learned, it should ideally be accompanied by semantically-congruent representational gesture.

Importantly, not all representational gestures affect comprehension similarly. For example, representational gestures that are semantically-incongruent with lexical affiliates (i.e., associated words or phrases) disrupt comprehension even more than the absence of gesture (Kelly et al., 2015; Dargue and Sweller, 2018). Moreover, representational gestures frequently produced in conjunction with lexical affiliates (e.g., holding up one finger to simulate *first place*) benefit comprehension more than representational gestures infrequently produced in conjunction with the same lexical affiliates (e.g., outlining a ribbon with ones hands to simulate *first place*; Dargue and Sweller, 2018). Although both gestures convey the concept of *first place,* frequently-produced representational gestures are thought to enhance comprehension because such gestures are more semantically-related to co-occuring speech—and are therefore more easily processed—than infrequently-produced representational gestures (Woodall and Folger, 1981). By examining differences in language processing resulting from representational gestures that are related to co-occuring speech to varying degrees, these findings emphasize the importance of semantic congruency between gesture and speech in lightening cognitive load and thereby enhancing language comprehension.

Although extant research has examined the semantic relationship between representational gesture and speech, it is currently unknown whether the learning implications of the Integrated Systems Hypothesis (i.e., increased comprehension) extend to the semantic relationship between representational gesture and text. Similar to how speech conveys information acoustically, text conveys information orthographically and is therefore a component of mental representations of language (Özyürek, 2002; Melinger and Levelt, 2005). Unlike speech, however, text is comprehended within the visual modality; therefore, it must be processed sequentially with gesture. To our knowledge, no published research to date has investigated how the semantic relationship between representational gesture and text is represented, despite that text is the orthographic equivalent of speech.

Understanding whether gesture and text are integrated similarly to gesture and speech is crucial in furthering the understanding of gesture's impact on language learning. When novel vocabulary is learned in instructional settings, words are often displayed in orthographic, as well as spoken, form. For example, a student may see a vocabulary word displayed on the white board or screen before seeing a gesture depicting what that word means. In order to determine whether representational gesture affects text comprehension in a similar manner to speech comprehension, it is first necessary to understand whether the semantic congruency of words presented via text with representational gestures is represented similarly to the semantic congruency of words presented via speech with representational gestures. Thus, the primary purpose of the present research was to compare how semantic congruency is represented, as evidenced by ratings, when representational gesture occurs with text vs. speech.

A secondary purpose of the present research was to provide an open normed database of semantically-congruent and semantically-incongruent gesture-word pairs in both text and speech for use in future research. Although a number of previous experiments have manipulated the semantic congruency of representational gestures and words relative to one another (Kelly et al., 2004, 2009, 2015; Özyürek et al., 2007; Straube et al., 2009; Dargue and Sweller, 2018), in most cases, the semantic congruency of gesture-word pairs was not normed. In light of this lack of norming data and evidence that the semantic relationship between representational gesture and lexical affiliates may fall along a continuum (Kelly et al., 2010; Dargue and Sweller, 2018), the degree of item-level variation within semantic congruence categories should be taken into consideration in future research. To minimize within-category variation in the present research, we constructed semantically-congruent gesture-word pairs from representational gestures and lexical affiliates (words) that they were consistently associated with, and we constructed semantically-incongruent gesture-word pairs from representational gestures and lexical affiliates with dissimilar, non-confusable forms and meanings.

To achieve our research objectives, we collected and compared semantic similarity ratings for representational gestures paired with semantically-congruent and semantically-incongruent words as speech and text. We predicted that semantically-congruent gesture-word pairs would be evaluated as highly semantically-related regardless of whether words were presented via text or speech, and that semantically-incongruent gesture-word pairs would be evaluated as highly semantically-unrelated regardless of whether words were presented via text or speech. These results would provide evidence that the semantic relationship between representational gesture and text, as evidenced by semantic congruency ratings, is represented similarly to the semantic relationship between representational gesture and speech.

METHOD

Participants

Two studies—a gesture-text and a gesture-speech study—were conducted via the internet with separate groups of participants. Sixty-nine participants were recruited for the gesture-text study, and seventy-one participants were recruited for the gesture-speech study. Participants ($n = 140$) were recruited from a large public university in the Southeastern United States in return for partial course credit. All participants were 18–35-year-old native English speakers who reported normal hearing and normal or corrected-to-normal vision and no speech, language, or learning impairments[1]. All participants provided consent to participate, and all experimental procedures were approved by the university's institutional review board.

Materials

All of the materials used in this research are publicly available via the Open Science Framework and can be accessed via the following link: https://osf.io/z5s3d/. Ninety-six English action verbs and 96 videos of representational gestures depicting their meanings (see **Supplementary Appendix A**) were used in the gesture-text and gesture-speech studies. Verbs were selected considering their frequency of use and the degree to which their meanings could be transparently conveyed via representational gesture. Representational gesture videos featured a woman silently enacting word meanings using the hands, body, and facial expressions. To ensure that spoken words did not differ in qualities such as affect, speed, or pitch based on their meanings, they were generated using the Microsoft Zira Desktop (Balabolka) text-to-speech synthesizer [English (United States, Female)].

Using these representational gesture and word stimuli, two types of gesture-word pairs were constructed for use in this study: Pairs consisting of gestures and words matched in meaning (semantically-congruent pairs), and pairs consisting of gestures and words mismatched in meaning (semantically-incongruent pairs; see **Supplementary Appendix B**). Construction of semantically-congruent and semantically-incongruent gesture-word pairs was based on data collected from a norming study in which 32 additional participants, who did not participate in the gesture-text or gesture-speech studies, selected the word best representing the action portrayed in each gesture video from among four alternatives. Based on this norming data, congruent gesture-word pairs were constructed from gestures reliably associated with their corresponding words, and incongruent gesture-word pairs were constructed from gestures and words with dissimilar, non-confusable forms and meanings.

Based on these gesture-word pairs, two lists were created for use in the gesture-text and gesture-speech studies. In these lists, gesture-word pairs were randomly divided in half and assigned to each congruency condition, such that gesture-word pairs that were semantically-congruent in one list were semantically-incongruent in the other list and vice versa. Order of presentation was randomized per participant such that semantically-congruent and semantically-incongruent gesture-word pairs were randomly interleaved in each study.

Procedure

Participants were provided with an anonymized link to either the gesture-text or gesture-speech study. Upon following this link to initiate their respective studies, which were administered using the Qualtrics platform, participants were randomly assigned to one of the two lists of gesture-word pairs divided by semantic congruency described above.

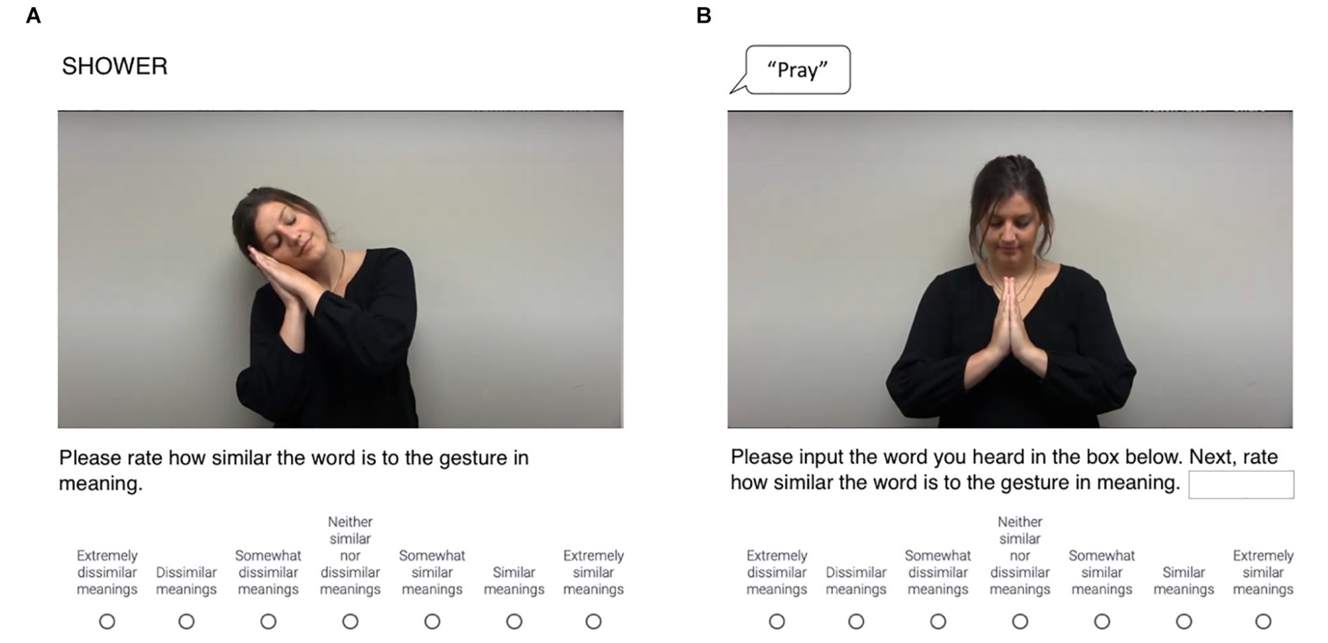
In the gesture-text study, participants viewed words as text and subsequently watched video clips of representational gestures that were either semantically-congruent or semantically-incongruent with them. Participants then rated the similarity of the meanings of these words and gestures using a 7-point Likert scale ranging from 1 (extremely dissimilar) to 7 (extremely similar; see **Figure 1A**). In the gesture-speech study, participants played audio clips of spoken words and subsequently played video clips of representational gestures that were either semantically-congruent or semantically-incongruent with them. For each item, participants then typed the spoken word that they heard into a text box to ensure that they understood it correctly and subsequently rated the semantic similarity of the meaning of that word and the gesture that they had viewed using the same scale as the gesture-text study (see **Figure 1B**).

RESULTS

All of the data and analysis scripts used in this research are publicly available via the Open Science Framework and can be accessed via the following link: https://osf.io/z5s3d/. **Table 1** displays frequency counts of semantic relatedness ratings by language modality and semantic congruency. Prior to analysis, words typed incorrectly in the gesture-speech study (22% of observations) were excluded. Semantic relatedness ratings for gesture-word pairs were then analyzed using a linear mixed effects model that included fixed effects of language modality and semantic congruency as well as random effects of participant and item with random slopes of congruency by participant, as follows:

$$\text{lmer(rating} \sim \text{modality} \times \text{congruency)}$$
$$+ (1 + \text{congruency}|\text{participant}) + (1|\text{item})$$

[1] These criteria were used for pre-screening. No demographic data were collected.



Please rate how similar the word is to the gesture in meaning.

Extremely dissimilar meanings	Dissimilar meanings	Somewhat dissimilar meanings	Neither similar nor dissimilar meanings	Somewhat similar meanings	Similar meanings	Extremely similar meanings
○	○	○	○	○	○	○

Please input the word you heard in the box below. Next, rate how similar the word is to the gesture in meaning. □

Extremely dissimilar meanings	Dissimilar meanings	Somewhat dissimilar meanings	Neither similar nor dissimilar meanings	Somewhat similar meanings	Similar meanings	Extremely similar meanings
○	○	○	○	○	○	○

FIGURE 1 | Schematic of **(A)** item featuring semantically-incongruent gesture-word pair from gesture-text task (Sleep—Shower); **(B)** item featuring semantically-congruent gesture-word pair from gesture-speech task (Pray—Pray).

TABLE 1 | Frequency of semantic relatedness ratings for gesture-word pairs by language modality and semantic congruency.

		Rating						
Language modality	**Semantic congruency**	**Extremely dissimilar**	**Dissimilar**	**Somewhat dissimilar**	**Neither similar nor dissimilar**	**Somewhat similar**	**Similar**	**Extremely similar**
		1	**2**	**3**	**4**	**5**	**6**	**7**
Speech	Congruent	13	24	39	52	351	711	1181
	Incongruent	1139	453	103	101	137	60	22
Text	Congruent	41	60	93	93	349	794	1161
	Incongruent	1295	640	180	121	211	105	39

This model was fit with Laplace estimation using the lmer() function of the lme4 package (Bates et al., 2015) in the R statistical programming language. Weighted mean-centered (Helmert) contrast coding was applied to all fixed effects (modality: speech = −0.54, text: 0.46; congruency: congruent = −0.48, incongruent: 0.52) using the psycholing package (Fraundorf, 2017) to obtain estimates analogous to those that would be obtained via ANOVA.

Table 2 displays parameter estimates for the model, and **Figure 2** displays semantic relatedness ratings assigned to gesture-word pairs by semantic congruency and language modality. We observed a significant main effect of semantic congruency, indicating that the meanings of gestures and words in semantically-congruent pairs ($M = 6.07$; SD = 1.23) were rated as more similar than the meanings of gestures and words in semantically-incongruent pairs ($M = 2.07$; SD = 1.54). By contrast, the main effect of language modality failed to reach significance, indicating that the meanings of paired gestures and words were rated similarly regardless of whether words were presented via speech ($M = 4.25$; SD = 2.46) or text ($M = 4.05$; SD = 2.41). Although we observed a non-significant trend toward an interaction between semantic congruency and language modality, simple main effect analyses by language modality revealed that the meanings of gestures and words in semantically-congruent pairs were rated as more similar than the meanings of gestures and words in semantically-incongruent pairs both when words were presented via speech ($M_{\text{congruent}} = 6.19$, $\text{SD}_{\text{congruent}} = 1.07$; $M_{\text{incongruent}} = 1.96$, $\text{SD}_{\text{incongruent}} = 1.48$; $B = -4.27$, SE = 0.12, $t = -35.21$, and $p < 0.001$), as well as text ($M_{\text{congruent}} = 5.96$, $\text{SD}_{\text{congruent}} = 1.35$; $M_{\text{incongruent}} = 2.14$, $\text{SD}_{\text{incongruent}} = 1.58$; $B = -3.81$, SE = 0.11, $t = -35.42$, and $p < 0.001$). Likewise, simple main effect analyses by semantic congruency revealed that the meanings of gestures and words presented via speech and text were rated similarly regardless of whether their meanings were congruent ($B = 0.08$, SE = 0.09, $t = 0.95$, and $p = 0.34$) or incongruent ($B = -0.08$, $SE = 0.10$,

TABLE 2 | Fixed effect estimates (Top) and variance estimates (Bottom) for multi-level linear model of semantic relatedness ratings of gesture-word pairs (observations = 9568).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept	4.14	00.07	62.89	<0.001***
Semantic congruency	−3.86	0.17	−22.44	<0.001***
Language modality	−0.03	0.06	−0.60	0.55
Semantic congruency × language modality	−0.25	0.14	−1.80	$0.07^{\dagger}$
Random effect				s^2
Participant				0.34
Participant × semantic congruency				1.28
Item				0.75

$^{\dagger}p < 0.1$; $^{***}p < 0.001$.

$t = -0.80$, and $p = 0.43$). Together, these findings indicate that the semantic relatedness of gesture-word pairs is not affected by language modality (speech vs. text).

DISCUSSION

The present research investigated how semantic congruency is evaluated when representational gesture is paired with lexical affiliates (words) conveyed via text vs. speech. Consistent with our hypothesis that semantic congruency ratings for semantically-congruent and semantically-incongruent gesture-word pairs would not differ based on whether words were presented via text or speech, the results indicate that ratings differed by semantic congruency but not language modality. These findings provide evidence that the semantic relationship between representational gestures and their lexical affiliates is evaluated similarly regardless of whether lexical affiliates are conveyed via the spoken or written modality.

These preliminary findings can be leveraged to further investigate whether semantically-congruent representational gesture is as beneficial to text comprehension as it is to speech comprehension, as posited by the Integrated Systems Hypothesis. The Integrated Systems Hypothesis postulates that representational gesture and speech are obligatorily and bi-directionally processed to enhance language comprehension (Kelly et al., 2010). Although integration was not directly investigated in the current study using online measures, similar congruency ratings for gesture-word pairs presented via speech and text indicate that the semantic relationships between representational gestures and their lexical affiliates are evaluated similarly during both speech and text processing. Therefore, we hypothesize that representational gesture may be obligatorily and bi-directionally integrated with text, similar to speech, during language processing.

Future research should further probe the relationship between gesture-speech and gesture-text processing using additional methods. Online behavioral measures such as reaction time, eye-tracking, and mouse-tracking may provide further evidence of whether representational gesture is integrated with text similarly to how it is integrated with

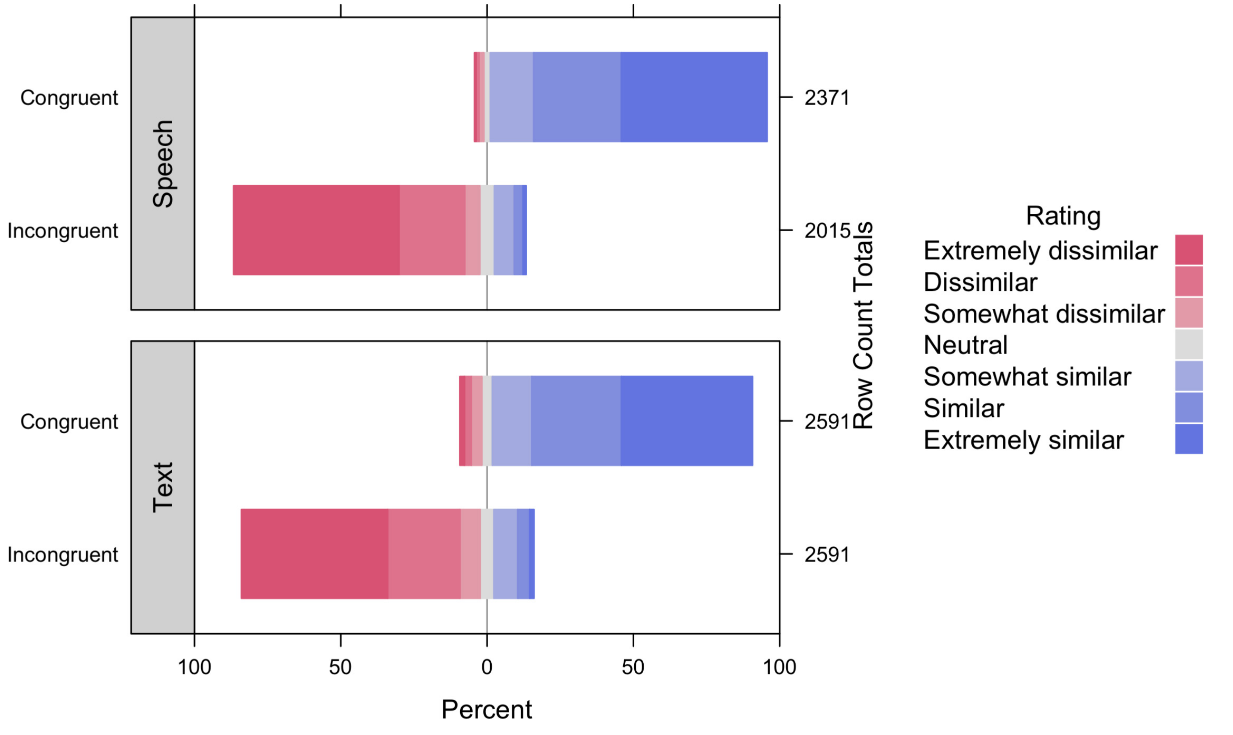


FIGURE 2 | Percentage of semantic relatedness ratings assigned to gesture-word pairs by semantic congruency and language modality.

speech during language processing. Moreover, building on previous evidence that gesture and speech are processed simultaneously during language comprehension (Özyürek et al., 2007), cognitive neuroscience methods with high temporal resolution, such as event-related potentials, can illuminate whether gesture and text are integrated simultaneously during language comprehension, similar to gesture and speech. Finally, cognitive neuroscience methods with high spatial resolution, such as functional magnetic resonance imaging, can be leveraged to further investigate the extent to which functional activity subserving gesture-text integration overlaps with functional activity subserving gesture-speech integration (Willems et al., 2007).

Although cognitive load was not assessed directly in the current research, the results provide preliminary evidence supporting further investigation into whether semantically-congruent representational gesture accompanied by speech and semantically-congruent representational gesture accompanied by text reduces cognitive load, benefiting language comprehension (Kelly et al., 2009, 2010). Cognitive Load Theory indicates that splitting cognitive resources between the verbal and visuospatial representational systems may decrease the cognitive demands of language processing, thereby improving comprehension (Clark and Paivio, 1991). Future research should directly investigate the effect of semantically-congruent representational gesture on cognitive load during text comprehension by measuring comprehenders' cognitive load while processing text accompanied by semantically-congruent representational gesture. Based on the findings of the current research, we hypothesize that representational gesture semantically related to sequentially-occurring language in both the spoken and written modalities may split comprehenders' cognitive load between the verbal and visuospatial representational systems, thereby enriching representations of language during comprehension.

In addition to providing insight into the similarity of semantic congruency between representational gesture and text vs. representational gesture and speech, the present research provides an open database of semantically-congruent and semantically-incongruent gesture-word pairs normed for semantic relatedness in both text and speech. These stimuli and ratings will be useful for future research investigating how semantic congruency of representational gesture affects processing of spoken vs. read language, particularly with respect to controlling for item-level variability within semantic congruence categories. Thus, we hope that future research will utilize these materials to further illuminate the cognitive and neural mechanisms of gesture-language integration.

In sum, the results of the present research indicate that the semantic relationship between representational gesture and text is evaluated similarly to the semantic relationship between representational gesture and speech. Thus, these results provide preliminary evidence for future research to examine whether language processing—and learning and memory—may be enhanced not only by semantically-congruent representational gesture occurring with speech, but also with text. In particular, these results provide preliminary evidence in support of an investigation into whether language, regardless of the modality it is presented in (i.e., spoken or orthographic), is influenced by the semantic congruency of representational gesture, providing important insight into how the relationship between gesture and language is represented in the minds of comprehenders. The results of the current work may have important educational implications as vocabulary words are sometimes orthographically displayed and accompanied by representational gestures. Finally, this work provides an open source database of stimuli and ratings that can be used to investigate how semantically-congruent and semantically-incongruent representational gestures affect spoken and written language comprehension.

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: Open Science Framework: https://osf.io/z5s3d/

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Alabama Institutional Review Board. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SSH-B developed gesture and word stimuli, oversaw implementation and data collection for norming and gesture-text studies, drafted and edited introduction and discussion, and created **Figure 1**. LMM conceptualized research, analyzed data, drafted results, edited entire manuscript, and created tables and **Figure 2**. RB oversaw implementation and data collection for gesture-speech study, drafted and edited methods, and created **Supplementary Appendices A,B**. All authors contributed to the article and approved the submitted version.

FUNDING

The writing of this manuscript was funded by a Taube Scholarship awarded to SSH-B.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.575991/full#supplementary-material

REFERENCES

Bates, D., Maechler, M., and Bolker, B. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.

Clark, J. M., and Paivio, A. (1991). Dual coding theory and education. *Educ. Psychol. Rev.* 3, 149–210. doi: 10.1007/bf01320076

Dargue, N., and Sweller, N. (2018). Donald Duck's garden: the effects of observing iconic reinforcing and contradictory gestures on narrative comprehension. *J. Exp. Child Psychol.* 175, 96–107. doi: 10.1016/j.jecp.2018.06.004

Drijvers, L., and Özyürek, A. (2017). Visual context enhanced: the joint contribution of iconic gestures and visible speech to degraded speech comprehension. *J. Speech Lang. Hear. Res.* 60, 212–222. doi: 10.1044/2016_JSLHR-H-16-0101

Drijvers, L., and Özyürek, A. (2020). Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. *Lang. Speech* 63, 209–220. doi: 10.1177/0023830919831311

Fraundorf, S. (2017). *Psycholing: R Functions for Common Psycholinguistic and Cognitive Designs. R Package Version 0.5.2.*

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol. Bull.* 137, 297–315. doi: 10.1037/a0022128

Kelly, S. D., Healey, M., Özyürek, A., and Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychon. Bull. Rev.* 22, 517–523. doi: 10.3758/s13423-014-0681-7

Kelly, S. D., Kravitz, C., and Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* 89, 253–260. doi: 10.1016/S0093-934X(03)00335-3

Kelly, S. D., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to world learning in a foreign language. *Lang. Cogn. Process.* 24, 313–334. doi: 10.1080/01690960802365567

Kelly, S. D., Özyürek, A., and Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327

Kendon, A. (2000). "Language and gesture: unity or duality?," in *Language and Gesture*, ed. D. McNeill (Cambridge, MA: Cambridge University Press), 47–63. doi: 10.1017/cbo9780511620850.004

Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3

McNeill, D. (1992). *Hand and Mind*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

Melinger, A., and Levelt, W. J. M. (2005). Gesture and the communicative intention of the speaker. *Gesture* 4, 119–141. doi: 10.1075/gest.4.2.02mel

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *J. Mem. Lang.* 46, 688–704. doi: 10.1006/jmla.2001.2826

Özyürek, A., Willems, R. M., Kita, S., and Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* 19, 605–617. doi: 10.1162/jocn.2007.19.4.605

So, W. C., Sim Chen-Hui, C., and Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: is meaning in gesture important for memory recall? *Lang. Cogn. Process.* 27, 665–681. doi: 10.1080/01690965.2011.573220

Straube, B., Green, A., Weis, S., Chatterjee, A., and Kircher, T. (2009). Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *J. Cogn. Neurosci.* 21, 821–836. doi: 10.1162/jocn.2009.21053

Willems, R. M., Özyürek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Woodall, W. G., and Folger, J. P. (1981). Encoding specificity and nonverbal cue context: an expansion of episodic memory research. *Commun. Monographs* 48, 39–53. doi: 10.1080/03637758109376046

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.

REVIEW
published: 05 November 2020
doi: 10.3389/fpsyg.2020.573555

Gesture Use and Processing: A Review on Individual Differences in Cognitive Resources

Demet Özer and Tilbe Göksun*

Department of Psychology, Koç University, Istanbul, Turkey

Speakers use spontaneous hand gestures as they speak and think. These gestures serve many functions for speakers who produce them as well as for listeners who observe them. To date, studies in the gesture literature mostly focused on group-comparisons or the external sources of variation to examine when people use, process, and benefit from using and observing gestures. However, there are also internal sources of variation in gesture use and processing. People differ in how frequently they use gestures, how salient their gestures are, for what purposes they produce gestures, and how much they benefit from using and seeing gestures during comprehension and learning depending on their cognitive dispositions. This review addresses how individual differences in different cognitive skills relate to how people employ gestures in production and comprehension across different ages (from infancy through adulthood to healthy aging) from a functionalist perspective. We conclude that speakers and listeners can use gestures as a compensation tool during communication and thinking that interacts with individuals' cognitive dispositions.

Keywords: individual differences, gesture production, gesture processing, cognitive resources, functions of gestures

OPEN ACCESS

Edited by:
Naomi Sweller,
Macquarie University, Australia

Reviewed by:
Ferdinand Binkofski,
RWTH Aachen University, Germany
Spencer D. Kelly,
Colgate University, United States

****Correspondence:***
Demet Özer
dozer@ku.edu.tr

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

Received: *17 June 2020*
Accepted: *29 September 2020*
Published: *05 November 2020*

Citation:
Özer D and Göksun T (2020) Gesture Use and Processing: A Review on Individual Differences in Cognitive Resources. Front. Psychol. 11:573555. doi: 10.3389/fpsyg.2020.573555

INTRODUCTION

Human language occurs in a face-to-face interactional setting with the exchange of multiple multimodal cues such as eye-gaze, lip movements, body posture, and hand gestures. In this review paper, we focus on one of these multimodal cues: iconic hand gestures (henceforth, gestures) that represent objects, events, and actions. Speakers use an abundant number of gestures as they speak or think. These gestures serve many functions for speakers who produce them and for listeners who observe them (Goldin-Meadow et al., 2001; McNeill, 2005; Özyürek, 2014; Kita et al., 2017; Novack and Goldin-Meadow, 2017; Dargue et al., 2019). Although gesture and speech express meaning in a coordinated and integrated manner, gesturing is not mandatory for communication and, hence, shows variation across situations and individuals (Kita and Özyürek, 2003; Kendon, 2004; McNeill, 2005; Streeck, 2009). Speakers differ in how frequently they use gestures, how salient their gestures are, and how much they benefit from *using* gestures during encoding and learning. On the other hand, listeners also differ in how much they attend to the speaker's gestures and benefit from *observing* gestures during comprehension and learning. The current paper will discuss individual differences in gesture use and processing.

There is individual variation in all human traits. People exhibit individual differences in cognitive abilities such as working memory (WM) capacity, attention, speech production, and processing as well as language acquisition (e.g., Daneman and Green, 1986; Just and Carpenter, 1992; Bates et al., 1995; Kane and Engle, 2002; Broadway and Engle, 2011; Huettig and Janse, 2016; Kidd et al., 2018). Current theories in cognitive science have not fully accounted for the existence as well as the causes of these individual differences for scientific gain (Underwood, 1975; Vogel and Awh, 2008). Most of the earlier studies in the gesture literature disregarded the variation among individuals and focused on group comparisons based on age (e.g., Feyereisen and Havard, 1999; Colletta et al., 2010; Austin and Sweller, 2014; Özer et al., 2017), sex (e.g., Özçalışkan and Goldin-Meadow, 2010), neuropsychological impairments (e.g., Cleary et al., 2011; Göksun et al., 2013b, 2015; Akbıyık et al., 2018; Akhavan et al., 2018; Hilverman et al., 2018; Özer et al., 2019; see Clough and Duff, 2020 for a review), culture, and the native status of the speakers and the listeners (i.e., bilinguals vs. monolinguals; e.g., Goldin-Meadow and Saltzman, 2000; Mayberry and Nicoladis, 2000; Pika et al., 2006; Kita, 2009; Nicoladis et al., 2009; Gullberg, 2010; Smithson et al., 2011; Kim and Lausberg, 2018; Azar et al., 2019, 2020) to understand how human multimodal language faculty operates at a general level. The gesture theories and current experimental practices in the gesture literature mostly downplayed the significance of individual differences and treated them as error variance. These studies create an illusionary and incorrect assumption that gesturing and the cognitive and communicative benefits of using and seeing gestures are invariant across people. However, using and observing gestures show not only across-group but also within-group variation (e.g., Hostetter and Alibali, 2007, 2011; Chu et al., 2014; Wu and Coulson, 2014a,b; Dargue et al., 2019; Özer et al., 2019; Özer and Göksun, 2020). What drives this variation?

There are external and internal sources of variation in gesture use and processing. The external sources of variation could be speech content (e.g., spatial vs. non-spatial topics; Rauscher et al., 1996; Feyereisen and Havard, 1999; Lausberg and Kita, 2003; Alibali, 2005; Hostetter, 2011), communicative context (e.g., the visibility between interlocutors, communicative intention, and audience design; Alibali et al., 2001; Trujillo et al., 2018; Schubotz et al., 2019), task difficulty and cognitive load levels (e.g., complex spatial tasks such as mental rotation; Wesp et al., 2001; Kita and Davies, 2009). There are also internal sources variation; even under the same external circumstances, people can behave differently. Insights into which mechanisms contribute to these individual differences just started to emerge (e.g., Hostetter and Alibali, 2007, 2011; Chu et al., 2014; Wu and Coulson, 2014a,b; Dargue et al., 2019; Aldugom et al., 2020; Kartalkanat and Göksun, 2020; Özer and Göksun, 2020).

Individual differences in personality characteristics, age, cognitive, and perceptual skills contribute to variation among individuals in terms of gesture use and processing (e.g., Vanetti and Allen, 1988; Cohen and Borsoi, 1996; Hostetter and Alibali, 2007, 2011; Wartenburger et al., 2010; Hostetter and Potthoff, 2012; Marstaller and Burianová, 2013; Göksun et al., 2013a; Chu et al., 2014; Gillespie et al., 2014; Wu and Coulson, 2014a,b; Pouw et al., 2016; Austin and Sweller, 2017, 2018; Eielts et al., 2018; Galati et al., 2018; Dargue and Sweller, 2020; Kartalkanat and Göksun, 2020; Özer and Göksun, 2020). However, most of the research on individual differences focused on gesture production, particularly on the cognitive correlates of variation in spontaneous gesture use and how much people benefit from using gestures during problem-solving and encoding of information. Research on individual variation in how listeners attend to speakers' gestures and benefit from observing gestures for comprehension and learning is limited (Wu and Coulson, 2014a,b; Aldugom et al., 2020; Özer and Göksun, 2020).

In the current review paper, we discuss individual differences in (1) *gesture use*: how frequently speakers use gestures during spontaneous speech and how much they benefit from using gestures during task solving and learning and (2) *gesture processing*: how listeners attend to and process speakers' gestures and how much they benefit from observing speakers' gestures for online comprehension or subsequent learning. We specifically focus on individual differences in cognitive and perceptual abilities (see Hostetter and Potthoff, 2012 for personality characteristics). This review has three highlights: (1) we attempt to bring a complete picture of individual differences in gesture by bridging production (i.e., using gestures) and comprehension (e.g., seeing gestures) fields. (2) We adopt a *functionalist* approach to discuss possible cognitive correlates of gesture use and processing. *Functionalist gesture theories* (as opposed to *mechanistic approaches* such as McNeill, 1992, 2005; Hostetter and Alibali, 2008) discuss *why* speakers use gestures and what *functions* gestures serve for speakers and listeners during communication and thinking (e.g., Kita and Özyürek, 2003; Pouw et al., 2014; Cook and Fenn, 2017; Kita et al., 2017; Novack and Goldin-Meadow, 2017). Theories asserting for what purposes speakers and listeners employ gestures might inform us about the possible cognitive correlates of individual differences in gesture use and processing. (3) We also take on a life-span developmental perspective, which covers how gesture use and processing differ in changing cognitive skills throughout the developmental trajectory (from childhood through adulthood to healthy aging).

The literature on how different populations across ages use and process gestures during communication and learning is quite rich. It is noteworthy that the current paper is not a comprehensive review of the general literature. Instead, we specifically focus on studies investigating individual differences in these processes. We first review the functions of gestures during communication and learning (section Functions of Gestures During Communication and Learning). Then, we address evidence on individual differences in gesture use (section Individual Differences in Gesture Production) and gesture processing (section Individual Differences in Gesture Processing) for children, young adults, and elderly adults. Last, we conclude the current state of the field and discuss areas that are open to further investigation (section Conclusion and Future Directions).

FUNCTIONS OF GESTURES DURING COMMUNICATION AND LEARNING

Several theories suggest how and why gestures occur during communication and thinking. Mechanistic theories mostly propose *how* gestures arise during communication and thinking (e.g., McNeill, 1992, 2005; Hostetter and Alibali, 2008, 2018). Functionalist theories, on the other hand, try to explain *why* we use gestures and the functions that gestures serve during communication and thinking, both for the speaker and the listener (e.g., Goldin-Meadow et al., 2001; Kita and Özyürek, 2003; Pouw et al., 2014; Cook and Fenn, 2017; Kita et al., 2017; Novack and Goldin-Meadow, 2017). The approach in this review will be from functionalist perspectives as they can give insight into which mechanisms might contribute to individual differences in gesture use and processing.

Gestures have several functions during communication and thinking. First, gestures affect communication between interlocutors. Speakers and listeners employ gestures for communicative purposes. Speakers produce gestures to communicate information, and listeners, in turn, benefit from these gestures to comprehend the to-be-communicated message (e.g., Beattie and Shovelton, 1999; Alibali et al., 2001; Holler and Stevens, 2007; Hostetter, 2011; Goldin-Meadow and Alibali, 2013). Speakers use gestures as an alternative channel of expression. Hence, both speakers and listeners employ gestures more in communicative challenges stemming from cognitive dispositions such as when a speaker is linguistically non-competent (e.g., bilinguals talking in their non-native language; Smithson et al., 2011; Gullberg, 2010) or has hearing impairments (Obermeier et al., 2012). The communicative function of gestures suggests that speakers and listeners with low communicative capacity (e.g., low linguistic proficiency, low semantic fluency, or the non-native status of the speaker and the listener) might employ and benefit from gestures more.

Second, gestures affect speakers' and listeners' cognitive processes. Gestures help activate, maintain, manipulate, and package visual, spatial, and motoric information for speaking and thinking (Kita et al., 2017). Gestures reduce cognitive load by keeping spatial-motoric information active in WM (Goldin-Meadow et al., 2001; Wesp et al., 2001; Morsella and Krauss, 2004; Ping and Goldin-Meadow, 2010; Cook et al., 2012; Marstaller and Burianová, 2013) and by projecting internal representations to an external space (e.g., Pouw et al., 2014). Producing gestures provides an external visual feedback that can be used to maintain or retrieve task-related visual-spatial information and, hence, reduces the cognitive load. Considering this, we expect that people might use gestures as a compensatory tool to manage their cognitive load. For example, people with lower visual-spatial cognitive capacity (e.g., lower visual-spatial WM capacity, lower general spatial skills assessed by mental rotation, and lower fluid intelligence assessed by Raven's Matrices) might use gestures more frequently to compensate for their limited resources when talking and thinking, especially about spatial information (e.g., Trafton et al., 2006; Göksun et al., 2013a; Chu et al., 2014; Galati et al., 2018). In a similar vein, speakers' gestures provide a stable visual representation for observers (i.e., listeners) and help listeners during comprehension and learning. People with lower cognitive resources might be in a greater need for external aids, and thus benefit more from seeing gestures (e.g., de Nooijer et al., 2013; Wu and Coulson, 2014a; Özer and Göksun, 2020).

Functional gesture theories assert that gestures help to convey information during communication and manage cognitive load during speaking, thinking, and learning (e.g., Kita et al., 2017; Novack and Goldin-Meadow, 2017). This suggests that gesture use and processing are sensitive to the cognitive dispositions of the speakers and the listeners. People might convey gestures to manage and compensate for their limited cognitive resources.

Mechanistic gesture theories, on the other hand, emphasize how people employ gestures. As opposed to functionalist theories, one of the first and most influential mechanistic accounts of gesture production (*The Growth Point Theory*, McNeill, 1992, 2005; McNeill and Duncan, 2000) posits that gestures do not compensate for thinking and speaking. According to this account, gesture and speech originate from a single representational system, where an utterance contains both linguistic and imagistic structures that cannot be separated. Speech stems from propositional linguistic representations and gestures stem from non-propositional imagistic representations and reflect visual, spatial, and motoric thinking (McNeill, 1992, 2005; Krauss et al., 2000). This account suggests that gestures are the manifestations of the imagistic component of the thought. Although mechanistic accounts would not be against the role gestures play for people to manage cognitive processes, they emphasize *how* people employ gestures rather than *why* they gesture.

In the following sections, we review evidence regarding how individual differences in cognitive domains relate with gesture use and processing from a functionalist account, mainly considering the *gesture-as-a-compensation-tool* view. That is, following the functionalist approach, we will illustrate the functions of gestures for speakers and listeners who use their cognitive resources differently. Gestures might not be used as a compensatory tool for every situation across different groups (e.g., So et al., 2009; Chui, 2011; de Ruiter et al., 2012); yet, the current state of the field supports the beneficial part of gestures for communication, thinking, and learning (e.g., Goldin-Meadow et al., 2001; Kita et al., 2017; Novack and Goldin-Meadow, 2017).

INDIVIDUAL DIFFERENCES IN GESTURE PRODUCTION

People from all ages show variation in terms of how frequently they use gestures, how salient their gestures are, and what types of gestures they use during spontaneous speech (e.g., Feyereisen and Havard, 1999; Richmond et al., 2003; Priesters and Mittelberg, 2013; Chu et al., 2014; Nagels et al., 2015; Schmalenbach et al., 2017; Arslan and Göksun, in press). People also differ in how much they benefit from using gestures during speaking, encoding, and subsequent learning (e.g., Goldin-Meadow et al., 2001; Ping and Goldin-Meadow, 2010; Galati et al., 2018). To date, studies mostly focused on two possible cognitive correlates: visual-spatial vs. verbal cognitive resources. We discuss how individual differences in visual-spatial and verbal cognitive capacities relate to gesture production in children, adults, and elderly adults.

Individual Differences in Gesture Production in Children

Babies start to use pointing gestures at around 12 months of age and iconic gestures at around 3 years of age (Iverson et al., 1999; Özçalışkan and Goldin-Meadow, 2005, 2010). Gestures open the way for the transition from prelinguistic to linguistic period, and gestures become increasingly intertwined with speech as children become older (e.g., Capirci et al., 2005; Capirci and Volterra, 2008; Liszkowski et al., 2008). Özçalışkan and Goldin-Meadow (2005) analyzed children's gestures at 14, 18, and 22 months of ages when children are interacting spontaneously with their mothers. They showed that children used more gestures as they got older. Moreover, there was a developmental shift toward the use of more supplementary gestures (e.g., saying "ride" and pointing at the bike) as opposed to reinforcing gestures (e.g., saying "bike" and pointing at the bike) by older children. Yet, there was no difference in the quality or the quantity of the maternal input across development, suggesting that changes in children's gestural behavior might reflect developmental changes in children's own cognitive processes. Then, individual differences in several cognitive processes might lead to variations in how and to what extent children use gestures in spontaneous speech. Children, even as early as 14 months of age, show individual variation in whether they use iconic gestures and how frequently they use them (e.g., Iverson et al., 1999; Özçalışkan and Goldin-Meadow, 2005). What drives these very early individual differences in gesture use? To date, the gesturing behavior of young children mostly focused on how individual differences in early gesture use predicted later language development (e.g., Rowe and Goldin-Meadow, 2009; Demir et al., 2015). Studies examining the precursors of these variations, on the other hand, primarily focused on how parental language input (speech and gesture) relates with children's spontaneous gesture production (e.g., Iverson et al., 2008; Rowe et al., 2008; Tamis-LeMonda et al., 2012). It is unknown which cognitive and perceptual abilities of these young children drive early individual differences in gesture production. Early socio-cognitive precursors of gesture use in infancy is an open area for further investigation.

How do children use gestures at later ages, such as during preschool and school-age? Children have not yet fully developed verbal skills as compared to young adults; thus, they might use gestures more during speaking as gestures provide an alternative channel of expression (e.g., Melinger and Levelt, 2004) and help facilitate speaking (Krauss et al., 2000). Indeed, studies report that preschool-aged children benefit more from gestures than older children and adults, especially when using complex language (e.g., Church et al., 2000; Austin and Sweller, 2014). Moreover, children in transitional stages (i.e., children who have the conceptual knowledge but not yet the skills to verbalize that knowledge) used more gestures to convey ideas compared to children who had necessary verbal resources to convey the same idea linguistically (e.g., Church and Goldin-Meadow, 1986; Perry et al., 1992). These gestures (so-called *gesture mismatches*) expressed non-redundant information that was not found in the accompanying speech. Children (ages 5–10) used more non-redundant speech-gesture combinations both at the clause and word levels compared to adults (Alibali et al., 2009). This is also evident in the expression of other linguistically challenging categories such as causal or spatial relations (e.g., Göksun et al., 2010; Austin and Sweller, 2018; Calero et al., 2019; Karadöller et al., 2019). Children used more gestures to convey additional information when they could not verbalize instruments of causal events (Göksun et al., 2010) or spatial relations such as left-right (Karadöller et al., 2019). For example, ambiguous spatial terms such as "*here*" can be complemented by gestures to specify the spatial relation (Karadöller et al., 2019). The multimodal discourse continues to develop during the school-age years. There is a developmental shift toward the use of a higher number of gestures per clause by 10-year-old children and adults than 6-year olds in narrative production tasks (e.g., Colletta et al., 2010; Alamillo et al., 2013).

Developmental studies suggest that children might use gestures as an alternative channel of expression to compensate for their limited linguistic proficiency (e.g., younger vs. older children or children vs. adults; Church et al., 2000; Alibali et al., 2009; Colletta et al., 2010). This is in line with bilingualism research showing that bilingual children speaking in their L2 used more gestures than monolinguals (e.g., Smithson et al., 2011; Wermelinger et al., 2020). Moreover, research on clinical populations with communication and language delays suggests that although there are delays in gesture production in the first 2 years, gesture might be used to compensate for communication and language difficulties at preschool and school ages by some children (Özçalışkan et al., 2013; LeBarton and Iverson, 2017). Children with language impairments (LI) used gestures at a higher rate and produced greater proportions of gestures that added unique information to the accompanying speech compared to typically developing (TD) peers, suggesting that children with LI employ gestures as an alternative channel of expression in the face of language difficulties (Evans et al., 2001; Blake et al., 2008; Iverson and Braddock, 2011; Mainela-Arnold et al., 2011, 2014).

Similar to children with LI, children with Down syndrome (DS) used more gesture-only expressions and expressed information uniquely in their gestures compared to TD children to compensate for spoken language delays (Stefanini et al., 2007; Dimitrova et al., 2016; Özçalışkan et al., 2017). Children with Williams syndrome (WS) also used more iconic gestures in a picture naming task compared to TD children to alleviate their word-finding difficulties (Bello et al., 2004). Yet, not all children with language delays benefit from gestures as a compensatory tool. Children with autism spectrum disorder (ASD) exhibit delays in gesture production that are apparent both in frequency and complexity (Colgan et al., 2006; Rozga et al., 2011; Watson et al., 2013; Dimitrova et al., 2016; Özçalışkan et al., 2016, 2017). Research shows that children with ASD used gestures to initiate and sustain joint attention and to compensate for speech limitations by supplementing speech to a lesser degree compared to TD peers, leading to negative consequences for learning and social interaction opportunities (Sowden et al., 2013; Watson et al., 2013; Mastrogiuseppe et al., 2015). Impairments in gesture production are more pronounced in ASD compared to other developmental delays such as DS

(Mastrogiuseppe et al., 2015), LI (Stone et al., 1997), and general intellectual delay (Mundy et al., 1990) and, thus, are considered to be a central component of problems in social interactions and delays in social development in ASD. Moreover, language delays not only affect children's gesture productions but also gestural input they receive from their caregivers, resulting in cascading consequences for language development.

Research suggests that children's language level affects caregivers' gestures to a greater extent when a child's language skills are limited (Iverson et al., 2006; Talbott et al., 2015; Dimitrova et al., 2016; Özçalışkan et al., 2017, 2018). For example, mothers of non-diagnosed high-risk ASD infants gestured more frequently compared to mothers of low-risk ASD infants (Talbott et al., 2015). The evidence on the compensatory use of gestures by children with language delays indicate that gesture is a tool that should be harnessed to support learning, especially for child clinical populations (LeBarton and Iverson, 2017). Gesture is also an early diagnostic tool to foresee persistent language delay, especially for children with unilateral brain lesions (Sauer et al., 2010; Özçalışkan et al., 2013). Although these studies suggest a link between early spoken language abilities and gesture production in children, the direct evidence on how individual differences in early receptive and expressive language skills relate with spontaneous gesture use within children with and without language delays is quite limited (Kartalkanat and Göksun, 2020; Wermelinger et al., 2020).

A growing body of literature shows that using gestures benefit children's subsequent memory and learning (e.g., Alibali and DiRusso, 1999; Wakefield et al., 2018). Do all children benefit similarly from using gestures? Post et al. (2013) showed that children who simultaneously produced and observed gestures when learning grammatical rules performed worse than children who only observed gestures. However, the adverse effects of gesturing on learning were only visible for children with lower verbal skills, suggesting that producing and observing gestures simultaneously might be too cognitively demanding, especially for children with lower verbal resources (Kalyuga, 2007). Nevertheless, it should be noted that this study tested the effects of using gestures on learning under high cognitive load. There is no direct evidence on how verbal skills relate to how much children benefit from using gestures when they are under average cognitive load (e.g., without observing gestures simultaneously).

Developmental studies mostly compared different age groups (e.g., children vs. adults or younger vs. older children; e.g., Colletta et al., 2010), bilinguals vs. monolinguals (e.g., Mayberry and Nicoladis, 2000; Smithson et al., 2011), and clinical vs. non-clinical groups (e.g., Bello et al., 2004; Dimitrova et al., 2016; LeBarton and Iverson, 2017). These studies suggest that children use gestures as a compensatory tool, and individual differences in verbal skills play a role in how much children use and benefit from using gestures during learning. Moreover, visual-spatial skills follow a protracted development, and children show individual variation in visual-spatial abilities (Newcombe et al., 2013). Given that gestures are visual-spatial entities and help activate, maintain, and manipulate visual-spatial information (Kita et al., 2017), individual differences in visual-spatial abilities during childhood might affect how much children use gestures and benefit from using gestures for learning. However, there is no direct evidence on how individual differences in verbal and visual-spatial skills relate to children's gesture use, which begs for future research.

Individual Differences in Gesture Production in Young Adults

Most of the research on individual differences in gesture production focused on young adults. Studies showed that young adults with lower cognitive capacities used more spontaneous gestures and benefited more from using gestures (e.g., Chu et al., 2014; Gillespie et al., 2014; but see Hostetter and Alibali, 2007), supporting the functionalist accounts (Goldin-Meadow et al., 2001; Marstaller and Burianová, 2013; Kita et al., 2017) and gesture's role as a compensation tool. The visual-spatial cognitive capacity is related to how much speakers employ gestures during speaking and thinking. People with lower visual and spatial WM capacities, mental rotation skills, and spatial conceptualization abilities (Kita and Davies, 2009) used more gestures compared to high-spatial ability individuals when explaining abstract phrases or social dilemmas (Chu et al., 2014). In a spatial gesture elicitation task, Göksun et al. (2013a) asked young adults to describe how they solved mental rotation problems and found that people with lower spatial abilities (lower mental rotation scores) used more gestures compared to people who had higher scores. However, low- and high-spatial ability individuals not only differed in the frequency of gestures but also in the type of gestures they used. People with low-spatial ability used more static gestures depicting objects (i.e., cubes or whole objects), whereas high-spatial ability individuals used more dynamic gestures to express motion, such as rotation or direction or static gestures referring to object pieces (e.g., the bottom part of the L shape). This finding is in line with a previous study showing that although lower- and higher-fluid intelligence individuals (as measured by Raven's matrices) used an equal number of gestures when describing how to solve geometric analogies, people with higher fluid intelligence used more gestures to express motion than people with lower fluid intelligence (Wartenburger et al., 2010; Sassenberg et al., 2011).

Verbal cognitive capacity is another predictor for how and to what extent speakers use gestures (e.g., Baxter et al., 1968; Hostetter and Alibali, 2007, 2011; Nagpal et al., 2011; Smithson and Nicoladis, 2013; Gillespie et al., 2014; for cf. see Frick-Horbury, 2002 and Chu et al., 2014). Young adults with lower verbal abilities such as lower verbal WM capacity, vocabulary size, and semantic fluency (i.e., phonological and lexical retrieval ability) used more gestures during spontaneous speech than individuals with higher verbal abilities (e.g., Hostetter and Alibali, 2007, 2011; Smithson and Nicoladis, 2013; Gillespie et al., 2014; but see Chu et al., 2014). These findings corroborate with bilingual research, showing that bilinguals used more gestures when talking in their L2 compared to L1 or monolinguals (e.g., Gullberg, 1998; Nagpal et al., 2011). Verbal WM also predicted gesture frequency similarly in bilinguals and monolinguals (Smithson and Nicoladis, 2013).

Is there an interaction between verbal and spatial skills in gesture use? Hostetter and Alibali (2007) showed a quadratic relationship between verbal resources and spontaneous gesture use. People with the lowest and highest verbal skills (i.e., phonemic fluency) gestured more than people with average verbal skills when they were retelling a cartoon story and describing how to wrap a package. Moreover, low verbal/high visual-spatial individuals produced the largest number of gestures and used more non-redundant gestures (Vanetti and Allen, 1988; Hostetter and Alibali, 2007, 2011). This might suggest that gestures are more helpful when speakers have spatial information in the non-propositional format in mind but are unable to lexicalize or to encode verbally (e.g., Graham and Heywood, 1975; Krauss and Hadar, 1999).

Young adults also show individual variation in how much they benefit from using gestures during task solving or subsequent memory and learning (e.g., Marstaller and Burianová, 2013). Young adults use many gestures when encoding information that facilitates their subsequent memory and learning, especially for visual and spatial information (e.g., Chu and Kita, 2011; So et al., 2015). However, using gestures is especially beneficial for people with lower cognitive capacity (e.g., Marstaller and Burianová, 2013; Pouw et al., 2016; Galati et al., 2018). People who used gestures when trying to learn new routes had better memory in a subsequent navigation task; however, this was only evident for people with a lower spatial perspective-taking ability (Galati et al., 2018). Moreover, gesturing benefited problem solving under higher cognitive load (e.g., dual-task paradigm; Marstaller and Burianová, 2013) and when internal cognitive resources are taxed or limited (e.g., Pouw et al., 2016).

Individual differences in verbal and visual-spatial skills affect how much young adult speakers use and benefit from producing gestures during speaking and problem-solving. Conforming the *gesture-as-a-compensation-tool* account, speakers employ gestures to compensate for lower verbal and spatial cognitive resources. However, we should be cautious about the generalizability of these findings as to the use of different cognitive measures, and gesture elicitation tasks (e.g., spatial vs. non-spatial abstract) might yield different results. Further research is needed to replicate these conclusions across contexts.

Individual Differences in Gesture Production in Healthy Aging

Evidence on spontaneous gesture use in healthy aging is minimal. Most of the research compared young and elderly adults and showed that spontaneous gesture production and gesture imitation is impaired in aged populations (e.g., Cohen and Borsoi, 1996; Dimeck et al., 1998; Feyereisen and Havard, 1999). Elderly adults used less representational gestures compared to young adults, whereas overall gesture frequency or the use of non-representational gestures (e.g., beat or conduit gestures) was comparable across two groups (Cohen and Borsoi, 1996; Glosser et al., 1998; Feyereisen and Havard, 1999; Arslan and Göksun, in press; for c.f. see Özer et al., 2017; Schubotz et al., 2019). This might be due to declining visual-spatial cognitive resources in aging. For example, mental imagery declines with aging (e.g., Dror and Kosslyn, 1994; Copeland and Radvansky, 2007; Andersen and Ni, 2008) and, indeed, individual differences in mental imagery, but not spatial WM capacity was associated with how frequently young and elderly individuals used spontaneous gestures, particularly for a spatial address description task (Arslan and Göksun, in press). Elderly individuals were also impaired in designing their multimodal utterances for their addressees (i.e., audience design; Schubotz et al., 2019). When narrating comic cartoons, young adults used fewer gestures when they knew that their addressee also watched the comic cartoon compared to when their addressee did not see the cartoon. However, elderly adults used an equal number of gestures in both cases.

We might expect that declining visual-spatial skills in aging would lead to higher use of gestures by older adults than in younger ones. However, gestures might be used as a compensatory tool only to manage cognitive load when the person has the necessary/intact resources. Most of the studies comparing younger vs. older adults tested individuals who are older than 60 years of age (e.g., Cohen and Borsoi, 1996), and it is unknown whether visual-spatial skills are severely impaired in this age group. Less is also known on the decline in which cognitive abilities in healthy aging leads to age-related impairments in gesture production (but see Arslan and Göksun, in press). It is important to note that this area is open to investigation and future research should study the decline in which cognitive resources lead to impaired gesturing in aging, whether the effects of aging on gesturing is similar for everyone, and which cognitive resources might play a protective role for the decline of gesture production. More research is also needed to examine whether elderly individuals benefit from using gestures as young adults and children do or producing gestures impose an extra cognitive burden to their already limited cognitive resources.

Moreover, we mainly focused on gesture use in healthy aging, yet, the line of research on how people with neurodegenerative disorders use gestures is informative as well (e.g., Cleary et al., 2011; Rousseaux et al., 2012; Klooster et al., 2015; Akhavan et al., 2018; Özer et al., 2019). People with different types of neurodegenerative disorders such as Alzheimer's disease, primary progressive aphasia, and Parkinson's disease are natural targets to study gesture in aged populations because the prevalence rates of these diseases consistently increase with age (e.g., Jorm et al., 1987; Brayne et al., 2006). For example, Klooster et al. (2015) showed that the beneficial effects of using and observing gestures on new learning in a Tower of Hanoi paradigm were absent in elderly patients with intact declarative memory, but impaired procedural memory as a consequence of Parkinson's disease. This suggests that the procedural memory system supports the ability of gestures to drive new learning. Thus, the decline in different memory systems in different neurodegenerative disorders that increase with age might lead to variation in how elderly adults benefit from gestures during learning. Future studies should test the cognitive correlates of impaired gesture use in different neuropsychological groups.

INDIVIDUAL DIFFERENCES IN GESTURE PROCESSING

Listeners are sensitive to speakers' gestures and benefit from observing these gestures during online language comprehension, encoding, and subsequent memory and learning (Holler et al., 2009; Kelly et al., 2010; Hostetter, 2011; Dargue et al., 2019). The facilitative effects of observing gestures are evidenced across children (e.g., Cook et al., 2008; Austin and Sweller, 2014, 2017; Macoun and Sweller, 2016; Vogt and Kauschke, 2017; Holler et al., 2018; Aussems and Kita, 2019; Dargue and Sweller, 2020; Kartalkanat and Göksun, 2020) and young adults (e.g., Beattie and Shovelton, 1999; Roth, 2001; Holle and Gunter, 2007; Kelly et al., 2008; Hostetter, 2011; Rueckert et al., 2017; Dargue and Sweller, 2020). Research regarding individual differences in how listeners attend to and process speakers' gestures and how much they benefit from observing gestures during comprehension and learning is quite limited, especially when compared to the literature on individual differences in gesture production (e.g., Post et al., 2013; Wu and Coulson, 2014a,b; Yeo and Tzeng, 2019; Özer and Göksun, 2020). In the next subsections, we review evidence on individual differences in gesture processing and its effects on comprehension and learning in children, young adults, and elderly adults. Then, we discuss several possible cognitive mechanisms that might yield individual differences in gesture processing, suggesting new venues for future research.

Individual Differences in Gesture Processing in Children

Electrophysiological studies showed that children start to process iconic gestures as semantic entities like words at around 18 months of age (Sheehan et al., 2007). Behaviorally, they start to comprehend iconic gestures representing entities at around 3 years of age (Stanfield et al., 2014) and iconic gestures representing events at around 4 years of age (Glasser et al., 2018). Studies showed that 3-year olds could not integrate speech and gesture, whereas 5-year old and adults did (e.g., Sekine and Kita, 2015; Sekine et al., 2015). Moreover, children starting from 6 years of age integrate speech and gesture in an online fashion comparable to adults (Dick et al., 2012; Sekine et al., in press). Demir-Lira et al. (2018) showed that gesture-speech integration recruits the same neural network as in adults. Yet, this was true only for children who were able to successfully integrate speech and gesture behaviorally. Then, what drives these individual differences in early gesture-speech integration ability?

Gesture-speech integration requires a global developmental shift. The precursors of gesture comprehension and gesture-speech integration are unknown. Gestures are visual-spatial entities and the processing, and the interpretation of gestures requires visual-spatial cognitive resources (e.g., Kelly and Goldsmith, 2004). Children with lower visual-spatial skills might have difficulty in processing and comprehending gestures compared to children with higher visual-spatial skills. The global development of executive attention and general WM capacity, on the other hand, might play a role in gesture-speech integration. For example, children with lower overall WM capacity might have difficulty in maintaining and integrating two different kinds of information simultaneously, especially in offline integration tasks (e.g., Demir-Lira et al., 2018). Cognitive predictors of individual differences in children's gesture comprehension and gesture-speech integration abilities require further attention.

What about individual differences in the beneficial effects of observing gestures for subsequent learning? Not all children benefit from visual aids such as diagrammatical illustrations when learning math (e.g., (Cooper et al., 2017). Indeed, observing gestures does not assist all children's comprehension of narratives or learning new skills (e.g., Church et al., 2004; van Wermeskerken et al., 2016; Yeo and Tzeng, 2019; Bohn et al., 2020; Kartalkanat and Göksun, 2020). Kartalkanat and Göksun (2020) found a positive relationship with verbal skills and the beneficial effects of observing gestures, preschoolers with higher expressive language ability benefited more from observing iconic gestures in the encoding of spatial events. Bohn et al. (2020) found that children benefited from observing gestures when learning novel skills (e.g., how to open a novel apparatus) as they became older. On the other hand, Demir et al. (2014) showed that children with pre- and perinatal unilateral brain injury (BI) and had difficulty in narrative production benefited more from observing gestures when retelling narratives compared to TD children (Demir et al., 2014). Moreover, children with specific language impairment (SLI) benefited more from observing gestures compared to TD and used the same gestures they observed when retelling the inferred meaning of the spoken messages (Kirk et al., 2011). The contradictory findings regarding the relation between verbal abilities and the beneficial effects of observing gestures between children with language impairments and children with intact language abilities pose a challenge. We might expect TD children with lower verbal abilities to benefit more from observing gestures as predicted by *gesture-as-a-compensation-tool* account; however, young children's limited verbal resources might be already consumed with processing speech, leaving few resources to process and benefit from external visual cues (i.e., gestures; Kalyuga, 2007). Again, gestures might help children manage cognitive load when they have fully developed verbal abilities. However, children with language impairments might employ gestures to compensate for their already-impaired spoken language abilities.

Individual differences in verbal (e.g., digit span task; Kartalkanat and Göksun, 2020), visual (e.g., visual patterns task; van Wermeskerken et al., 2016), or general WM capacity (e.g., operation span task; Yeo and Tzeng, 2019) did not predict how much children benefited from observing gestures for learning. There was hardly any variance in WM capacity in most of these studies (e.g., van Wermeskerken et al., 2016). This might obscure the otherwise possible effects of different WM capacities on the values of observing gestures in children. Additionally, how general spatial skills (e.g., mental rotation and mental imagery) relate to how much children benefit from observing gestures needs to be investigated in future research.

Individual Differences in Gesture Processing in Young Adults

Young adults also differ in how they process spontaneous co-speech gestures. Processing gestures require visual, spatial, and motoric cognitive resources (e.g., Kelly and Goldsmith, 2004; Wu and Coulson, 2014a). We expect people with higher visual-spatial abilities to process and comprehend gestures better. Indeed, Wu and Coulson (2014a) found that people with higher spatial WM (but not verbal WM) were better at processing co-speech gestures as they were more sensitive to speech-gesture mismatches (i.e., high-spatial individuals were affected more negatively when gesture and speech expressed incongruent information). Moreover, people who have larger spans for retaining and manipulating bodily configurations (i.e., motor movement span task assessing individuals' ability to retain body-centric motor information) comprehended gestures better (Wu and Coulson, 2014b). In a recent study, we asked how visual-spatial vs. verbal WM capacity relates to processing concurrent visual (i.e., gesture) and verbal (i.e., speech) information in a mismatch paradigm used initially by Kelly and colleagues in 2011 (Özer and Göksun, 2020). We demonstrated that listeners showed differential sensitivity in processing concurrent gestural vs. spoken information. Although gesture-speech mismatches hindered overall comprehension, how listeners got affected by mismatches in different modalities (gesture vs. speech mismatches) was dependent on the listeners' cognitive dispositions on visual-spatial vs. verbal resources. Observing mismatching visual information (i.e., gesture) imposes an additional visual-spatial cognitive load and people with higher spatial abilities were better at maintaining and processing two different and mismatching visual information due to their higher capacity. As a result, these individuals perform better when gestures expressed mismatching information compared to people with lower spatial abilities. People with higher verbal abilities, on the other hand, had better performance when speech expressed mismatching information compared to people with lower verbal abilities. These findings suggest that visual-spatial cognitive resources are critical for gesture processing and observing mismatching gestures increase visual-spatial cognitive load (e.g., Kelly and Goldsmith, 2004; Hostetter et al., 2018). Processing mismatching information in visual modality would be less demanding for people with larger visual-spatial cognitive resources.

What about individual differences in how much listeners benefit from observing gestures? Earlier studies are limited in suggesting how listeners integrate visual information with speech and use gestures to encode information either for online language comprehension or for subsequent learning. Research on how learners benefit from different multimedia materials (visual vs. verbal representations) might give us insight in this matter (Ausburn and Ausburn, 1978; Kirby et al., 1988; Koć-Januchta et al., 2017; Kiat and Belli, 2018; but see Kirschner, 2017). Individuals show variation in how they benefit from visual vs. verbal information (Kirby et al., 1988; Riding et al., 1995; Kozhevnikov et al., 2002; Mendelson and Thorson, 2004; Meneghetti et al., 2014; Alfred and Kraemer, 2017). For example, learners show variation in how they fixated to text vs. pictures when learning from multimedia resources (Koć-Januchta et al., 2017) and students with higher spatial abilities benefited more from the presence of 3D models when learning cell biology compared to students with lower spatial abilities (Huk, 2006). This suggests that listeners' cognitive dispositions might be related to how much they benefit from observing gestures vs. hearing speech. A very recent study directly tested how different WM capacities related to how much young adults benefited from observing gestures (Aldugom et al., 2020). Undergraduate students with higher visual WM capacity (i.e., visual patterns task) benefited more from observing gestures during math learning whereas verbal (i.e., sentence span task) and motoric (i.e., movement span task, Wu and Coulson, 2014b) WM capacities did not predict the beneficial effects of observing gestures (Aldugom et al., 2020). Although it is well-established in the literature that gestures facilitate listeners' comprehension and learning (see Özyürek, 2014 for review), evidence suggests that this is not a monolithic process. It is also possible that observing gestures do not always facilitate comprehension and learning. For example, observing gestures hurt learning phonetic distinctions at the syllable level within a word for English-speaking adults learning vowel length contrasts in Japanese (Kelly et al., 2014). However, as in the case with children (Kartalkanat and Göksun, 2020), learners' level of second-language proficiency might play a role for benefitting from gestures, which is another cognitive resource to be examined. Future studies should investigate the cognitive precursors of individual differences in the beneficial effects of observing gestures across different learning contexts (e.g., spatial vs. non-spatial) and different stages of language processing (e.g., phonological vs. semantic; Kelly et al., 2014).

Individual Differences in Gesture Processing in Healthy Aging

Few studies examined how elderly individuals process gestures and benefit from observing gestures (e.g., Thompson, 1995; Ska and Croisile, 1998; Montepare et al., 1999; Thompson and Guzman, 1999; Cocks et al., 2011). Elderly individuals are impaired in their comprehension of pantomimes and emotional gestures compared to young individuals (Ska and Croisile, 1998; Montepare et al., 1999). Moreover, elderly adults are impaired in integrating speech and gesture compared to young adults (Cocks et al., 2011). However, they performed equally when two cues are presented in isolation, suggesting that they might be impaired in gesture-speech integration with a preserved ability to process gestures. Indeed, elderly adults mostly relied on visible speech and did not benefit from observing gestures when recalling sentences (Thompson, 1995).

Although young adults benefited from visual aids (i.e., visible speech and gestures) under challenging listening conditions (i.e., dichotic shadowing task), older adults did not (Thompson and Guzman, 1999). The differences in the effects of observing gestures between younger and older adults might be related to the declining cognitive abilities associated with aging, mainly due to WM capacity as WM is required to maintain and manipulate different information. However, this has not been addressed directly.

Previous research suggests that elderly adults have difficulty in integrating visual (i.e., gesture) and verbal (i.e., speech) information compared to younger adults. It might be due to a decline in global cognitive skills such as executive attention and general WM capacity. Yet, it has not been directly tested. Future studies should compare younger vs. older adults with several cognitive measures to understand the cognitive architecture behind impaired gesture processing and gesture-speech integration in healthy aging.

CONCLUSION AND FUTURE DIRECTIONS

Speakers use gestures as they speak and think, and listeners, in turn, are sensitive to speakers' gestures. Gestures (both using by speakers and observing by listeners) have beneficial effects on language comprehension, problem-solving, encoding, and subsequent learning. Studies, to date, mostly focused on the role of different external factors (e.g., speech content and communicative context) on gestural behavior to answer *when* we use and benefit from gestures. However, it is also essential to ask *who* uses and benefits from gestures for *which* purposes. Research on the cognitive precursors of these individual differences in gesture use and processing has just started to emerge. Examining individual differences in gesture use and processing will help us uncover the cognitive architecture behind these processes and inform gesture research that is based on group data. The accounts that explain how and why gestures are employed should integrate individual differences research to have a full picture of when, why, and for whom gestures exhibit their supposed roles. This line of research is also informative for the development of educational programs incorporating the use of gestures by learners or teachers. The instructional programs should be tailored according to the cognitive dispositions and needs of the learners for optimal learning outcomes.

Most of the research on individual differences in the gesture literature examined gesture production in young adults. Studies on gesture use in children and elderly adults focused on group comparisons (i.e., comparing children at different ages, children vs. adults, younger vs. older adults, and clinical vs. non-clinical groups). Moreover, individual differences in gesture processing are limited compared to the production literature. In the current review paper, we (1) combined two lines of research: using gestures and observing gestures and (2) discussed the possible cognitive precursors of gesture use and processing in different age groups. We also highlighted the functions of producing and seeing gestures regarding their compensatory roles in speaking and thinking.

Gestures provide an alternative expression channel and assist speakers and listeners communicate (e.g., Alibali et al., 2001; Hostetter, 2011). Gestures also decrease speakers' and listeners' cognitive load by aiding them to activate, maintain, and manipulate visual-spatial information (e.g., Kita et al., 2017; Novack and Goldin-Meadow, 2017). Gestures help people manage cognitive load and are used as a compensatory tool. Listeners' and speakers' cognitive dispositions interact with this compensatory role of gestures, leading to individual differences in how much people benefit from using and seeing gestures for speaking, comprehension, task solving, and learning. As the *gesture-as-a-compensation-tool* account would argue, children and adults with lower cognitive resources use and benefit from using gestures more to manage cognitive load compared to people with higher cognitive resources (e.g., Church et al., 2000; Göksun et al., 2010; Marstaller and Burianová, 2013; Austin and Sweller, 2014; Chu et al., 2014; Gillespie et al., 2014; Galati et al., 2018). However, we suggest that gestures do not replace the impaired cognitive abilities; instead, gestures help manage cognitive load when cognitive resources are intact. Gestures might not compensate for already-impaired cognitive abilities. For example, people with aphasia use more gestures to compensate for impaired speech, but only when they have the intact conceptual knowledge of what they express (e.g., Göksun et al., 2013b, 2015). In a similar vein, there is a decrease in gesture production in healthy aging that might be due to impaired visual-spatial abilities such as mental imagery (e.g., Cohen and Borsoi, 1996; Arslan and Göksun, in press). There is also evidence of individual differences in gesture processing. Processing and comprehending gestures require visual-spatial cognitive resources (e.g., Kelly and Goldsmith, 2004; Hostetter et al., 2018). People with higher visual-spatial skills (or older children compared to younger children) process gestures better compared to people with lower visual-spatial skills (e.g., Wu and Coulson, 2014a,b; Özer and Göksun, 2020). In line with the *gesture-as-a-compensation-tool* account, we expect that people with lower cognitive resources (especially visual-spatial) would benefit more from observing external visual cues (i.e., gestures, but see Aldugom et al., 2020). However, research on how visual-spatial abilities relate to how much listeners benefit from observing gestures is inconclusive and begs for further investigation.

Although group-comparison studies are informative, future work should address within-group variation more, especially in children and in elderly adults. How different cognitive skills are associated with gesture production and processing should be tested directly across different conditions. The employment of different cognitive measures, gesture elicitation tasks, and learning contexts might yield different results, and these should be incorporated to have a full picture of whom for and when gestures are helpful. For example, the relationship between visual-spatial abilities and how frequently speakers use gestures and how much they benefit from using and observing gestures depend on the content of the information to be communicated or learned (Lausberg and Kita, 2003; Hostetter and Alibali, 2007; Chu et al., 2014; Arslan and Göksun, in press). The role of visual-spatial abilities in gesture use and processing might be more pronounced in spatial speech compared to non-spatial speech (e.g., Alibali, 2005; Arslan and Göksun, in press). Future work should also investigate how internal sources of variation (e.g., individual differences in several abilities) interact with external sources of variation (e.g., speech content and task difficulty).

One area open for future investigation is the cognitive predictors of gesture processing; that is, how listeners attend, process, and benefit from observing gestures. Studies on how visual, verbal, and motoric WM capacities are linked to individuals' processing of concurrent gesture vs. speech employed mismatch paradigms (Wu and Coulson, 2014a,b; Özer and Göksun, 2020). However, gesture mismatches are rare in natural communication, and we should investigate how different cognitive abilities relate to gesture processing in more ecologically valid paradigms. It is also unknown whether there are any individual differences in visual attention to gestures. Gestures are visual articulators and subject to visual processing. Although earlier research found that gestures can be processed peripherally and do not require direct visual attention (e.g., Gullberg and Holmqvist, 1999, 2006; Gullberg and Kita, 2009), recent evidence suggests that several factors might modulate how listeners allocate overt visual attention to gestures such as the comprehensibility of speech, and the native/non-native status of the listener (e.g., Drijvers et al., 2019). Future studies should address whether people with different visual-spatial vs. verbal abilities show differential overt visual attention to gestures and how this relates to individual differences in gesture processing (Wakefield et al., 2018). Above attending to and processing gestures, very little is known on whether and how individuals benefit from observing gestures during online language comprehension and learning across different learning contexts (Aldugom et al., 2020). We currently investigate how visual-spatial skills relate to how much listeners benefit from observing gestures when comprehending spatial relations between objects.

All studies reviewed above tested individual differences behaviorally. Electrophysiological and neuroimaging studies investigate the neural architecture of gesture use and processing (e.g., Kelly et al., 2004; Wu and Coulson, 2005; Willems et al., 2009). We might observe individual differences in neural data, that is, otherwise non-observable behaviorally (e.g., Demir-Lira et al., 2018). Future work should examine individual differences in the recruitment of different neural networks when using and observing gestures and how these differences in neural data relate to behavioral performance after considering individuals' cognitive skills.

The current review only focused on how individual differences in cognitive skills (mostly verbal and visual-spatial skills) relate to gesture use and processing. However, individual differences in other domains might also affect how people employ gestures. Individual differences in other domains such as personality (Hostetter and Potthoff, 2012) and other aspects of cognitive and perceptual skills such as selective attention, auditory processing, and the speed of multisensory processing should be tested (e.g., Schmalenbach et al., 2017). Moreover, it is also important to study the relation between gesture production and processing. How individual differences in spontaneous gesture use predict how people attend to and benefit from observing gestures or vice versa are unknown (Wakefield et al., 2013). Gesture processing might be affected by to what extent people themselves use gestures, and future studies should address the production-perception cycle and the mechanisms behind it.

In sum, gesture use and processing are not monolithic processes and show individual variation. Speakers and listeners can use gestures as a compensation tool during communication and thinking that interact with individuals' cognitive dispositions.

AUTHOR CONTRIBUTIONS

DÖ and TG conceived of the presented idea. DÖ drafted the manuscript. TG revised the manuscript critically for important intellectual content. Both the authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by TÜBA-GEBİP 2018 award given (Turkish Academy of Sciences Outstanding Young Scientist Award) and a James S. McDonnell Foundation Scholar Award (Grant no: 220020510) to TG.

ACKNOWLEDGMENTS

We would like to thank all members of the Language & Cognition Lab at Koç University for thought-provoking discussions and invaluable contributions to various stages of this article.

REFERENCES

Akbıyık, S., Karaduman, A., Göksun, T., and Chatterjee, A. (2018). The relationship between co-speech gesture production and macrolinguistic discourse abilities in people with focal brain injury. *Neuropsychologia* 117, 440–453. doi: 10.1016/j.neuropsychologia.2018.06.025

Akhavan, N., Göksun, T., and Nozari, N. (2018). Integrity and function of gestures in aphasia. *Aphasiology* 32, 1310–1335. doi: 10.1080/02687038.2017.1396573

Alamillo, A. R., Colletta, J. M., and Guidetti, M. (2013). Gesture and language in narratives and explanations: the effects of age and communicative activity on late multimodal discourse development. *J. Child Lang.* 40, 511–538. doi: 10.1017/S0305000912000062

Aldugom, M., Fenn, K., and Cook, S. W. (2020). Gesture during math instruction specifically benefits learners with high visuospatial working memory capacity. *Cogn. Res. Princ. Implic.* 5:27. doi: 10.1186/s41235-020-00215-8

Alfred, K. L., and Kraemer, D. J. (2017). Verbal and visual cognition: individual differences in the lab, in the brain, and in the classroom. *Dev. Neuropsychol.* 42, 507–520. doi: 10.1080/87565641.2017.1401075

Alibali, M. W. (2005). Gesture in spatial cognition: expressing, communicating, and thinking about spatial information. *Spat. Cogn. Comput.* 5, 307–331. doi: 10.1111/j.1756-8765.2010.01113.x

Alibali, M. W., and DiRusso, A. A. (1999). The function of gesture in learning to count: more than keeping track. *Cogn. Dev.* 14, 37–56. doi: 10.1016/S0885-2014(99)80017-3

Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., and Mainela-Arnold, E. (2009). Gesture-speech integration in narrative: are children less redundant than adults? *Gesture* 9, 290–311. doi: 10.1075/gest.9.3.02ali

Alibali, M. W., Heath, D. C., and Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *J. Mem. Lang.* 44, 169–188. doi: 10.1006/jmla.2000.2752

Andersen, G. J., and Ni, R. (2008). Aging and visual processing: declines in spatial not temporal integration. *Vis. Res.* 48, 109–118. doi: 10.1016/j.visres.2007.10.026

Arslan, B., and Göksun, T. (in press). Aging, working memory, and mental imagery: understanding gestural communication in younger and older adults. *Q. J. Exp. Psychol.* doi: 10.1177/1747021820944696

Ausburn, L. J., and Ausburn, F. B. (1978). Cognitive styles: some information and implications for instructional design. *Educ. Technol. Res. Dev.* 26, 337–354.

Aussems, S., and Kita, S. (2019). Seeing iconic gestures while encoding events facilitates children's memory of these events. *Child Dev.* 90, 1123–1137. doi: 10.1111/cdev.12988

Austin, E. E., and Sweller, N. (2014). Presentation and production: the role of gesture in spatial communication. *J. Exp. Child Psychol.* 122, 92–103. doi: 10.1016/j.jecp.2013.12.008

Austin, E. E., and Sweller, N. (2017). Getting to the elephants: gesture and preschoolers' comprehension of route direction information. *J. Exp. Child Psychol.* 163, 1–14. doi: 10.1016/j.jecp.2017.05.016

Austin, E. E., and Sweller, N. (2018). Gesturing along the way: adults' and preschoolers' communication of route direction information. *J. Nonverbal Behav.* 42, 199–220. doi: 10.1007/s10919-017-0271-2

Azar, Z., Backus, A., and Özyürek, A. (2019). General-and language-specific factors influence reference tracking in speech and gesture in discourse. *Discourse Process.* 56, 553–574. doi: 10.1080/0163853X.2018.1519368

Azar, Z., Backus, A., and Özyürek, A. (2020). Language contact does not drive gesture transfer: heritage speakers maintain language specific gesture patterns in each language. *Biling.: Lang. Cogn.* 23, 414–428. doi: 10.1017/S136672891900018X

Bates, E., Dale, P., and Thal, D. (1995). "Individual differences and their implications for theories of language development" in *Handbook of child language*. eds. P. Fletcher and B. MacWhinney (Oxford: Blackwell).

Baxter, J. C., Winters, E. P., and Hammer, R. E. (1968). Gestural behavior during a brief interview as a function of cognitive variables. *J. Pers. Soc. Psychol.* 8:303. doi: 10.1037/h0025597

Beattie, G., and Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica* 123, 1–30.

Bello, A., Capirci, O., and Volterra, V. (2004). Lexical production in children with Williams syndrome: spontaneous use of gesture in a naming task. *Neuropsychologia* 42, 201–213. doi: 10.1016/s0028-3932(03)00172-6

Blake, J., Myszczyszyn, D., Jokel, A., and Bebiroglu, N. (2008). Gestures accompanying speech in specifically language-impaired children and their timing with speech. *First Lang.* 28, 237–253. doi: 10.1177/0142723707087583

Bohn, M., Kordt, C., Braun, M., Call, J., and Tomasello, M. (2020). Learning novel skills from iconic gestures: a developmental and evolutionary perspective. *Psychol. Sci.* 31, 873–880. doi: 10.1177/0956797620921519 in press

Brayne, C., Gao, L., Dewey, M., Matthews, F. E., and Ageing Study Investigators (2006). Dementia before death in ageing societies—the promise of prevention and the reality. *PLoS Med.* 3:e397. doi: 10.1371/journal.pmed.0030397

Broadway, J. M., and Engle, R. W. (2011). Individual differences in working memory capacity and temporal discrimination. *PLoS One* 6:e25422. doi: 10.1371/journal.pone.0025422

Calero, C. I., Shalom, D. E., Spelke, E. S., and Sigman, M. (2019). Language, gesture, and judgment: children's paths to abstract geometry. *J. Exp. Child Psychol.* 177, 70–85. doi: 10.1016/j.jecp.2018.07.015

Capirci, O., Contaldo, A., Caselli, M. C., and Volterra, V. (2005). From action to language through gesture: a longitudinal perspective. *Gesture* 5, 155–177. doi: 10.1075/gest.5.1.12cap

Capirci, O., and Volterra, V. (2008). Gesture and speech: the emergence and development of a strong and changing partnership. *Gesture* 8, 22–44. doi: 10.1075/gest.8.1.04cap

Chu, M., and Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *J. Exp. Psychol. Gen.* 140, 102–116. doi: 10.1037/a0021790

Chu, M., Meyer, A., Foulkes, L., and Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *J. Exp. Psychol. Gen.* 143, 694–709. doi: 10.1037/a0033861

Chui, K. (2011). Do gestures compensate for the omission of motion expression in speech? *Chinese Lang. Discourse* 2, 153–167. doi: 10.1075/cld.2.2.01chu

Church, R. B., Ayman-Nolley, S., and Mahootian, S. (2004). The role of gesture in bilingual education: does gesture enhance learning? *Int. J. Biling. Educ. Biling.* 7, 303–319. doi: 10.1080/13670050408667815

Church, R. B., and Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23, 43–71. doi: 10.1016/0010-0277(86)90053-3

Church, R. B., Kelly, S. D., and Lynch, K. (2000). Immediate memory for mismatched speech and representational gesture across development. *J. Nonverbal Behav.* 24, 151–174. doi: 10.1023/A:1006610013873

Cleary, R. A., Poliakoff, E., Galpin, A., Dick, J. P., and Holler, J. (2011). An investigation of co-speech gesture production during action description in Parkinson's disease. *Parkinsonism Relat. Disord.* 17, 753–756. doi: 10.1016/j.parkreldis.2011.08.001

Clough, S., and Duff, M. C. (2020). The role of gesture in communication and cognition: implications for understanding and treating neurogenic communication disorders. *Front. Hum. Neurosci.* 14:323. doi: 10.3389/fnhum.2020.00323

Cocks, N., Morgan, G., and Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture* 11, 24–39. doi: 10.1075/gest.11.1.02coc

Cohen, R. L., and Borsoi, D. (1996). The role of gestures in description-communication: a cross-sectional study of aging. *J. Nonverbal Behav.* 20, 45–63.

Colgan, S. E., Lanter, E., McComish, C., Watson, L. R., Crais, E. R., and Baranek, G. T. (2006). Analysis of social interaction gestures in infants with autism. *Child Neuropsychol.* 12, 307–319. doi: 10.1080/09297040600701360

Colletta, J. M., Pellenq, C., and Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: evidence from French children and adults. *Speech Comm.* 52, 565–576. doi: 10.1016/j.specom.2010.02.009

Cook, S. W., and Fenn, K. M. (2017). "The function of gesture in learning and memory" in *Why gesture? How the hands function in speaking, thinking and communicating*. eds. R. B. Church, M. W. Alibali and S. D. Kelly (Amsterdam: John Benjamins Publishing Company), 129–153.

Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition* 106, 1047–1058. doi: 10.1016/j.cognition.2007.04.010

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Lang. Cogn. Process.* 27, 594–610. doi: 10.1080/01690965.2011.567074

Cooper, J. L., Sidney, P. G., and Alibali, M. W. (2017). Who benefits from diagrams and illustrations in math problems? Ability and attitudes matter. *Appl. Cogn. Psychol.* 32, 24–38. doi: 10.1002/acp.3371

Copeland, D. E., and Radvansky, G. A. (2007). Aging and integrating spatial mental models. *Psychol. Aging* 22, 569–579. doi: 10.1037/0882-7974.22.3.569

Daneman, M., and Green, I. (1986). Individual differences in comprehending and producing words in context. *J. Mem. Lang.* 25, 1–18.

Dargue, N., and Sweller, N. (2020). Learning stories through gesture: gesture's effects on child and adult narrative comprehension. *Educ. Psychol. Rev.* 32, 249–276. doi: 10.1007/s10648-019-09505-0

Dargue, N., Sweller, N., and Jones, M. P. (2019). When our hands help us understand: a meta-analysis into the effects of gesture on comprehension. *Psychol. Bull.* 145, 765–784. doi: 10.1037/bul0000202

de Nooijer, J. A., van Gog, T., Paas, F., and Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychol.* 144, 173–179. doi: 10.1016/j.actpsy.2013.05.013

de Ruiter, J. P., Bangerter, A., and Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: investigating the tradeoff hypothesis. *Top. Cogn. Sci.* 4, 232–248. doi: 10.1111/j.1756-8765.2012.01183.x

Demir, Ö. E., Fisher, J. A., Goldin-Meadow, S., and Levine, S. C. (2014). Narrative processing in typically developing children and children with early unilateral brain injury: seeing gesture matters. *Dev. Psychol.* 50, 815–828. doi: 10.1037/a0034322

Demir, Ö. E., Levine, S. C., and Goldin-Meadow, S. (2015). A tale of two hands: children's early gesture use in narrative production predicts later narrative structure in speech. *J. Child Lang.* 42, 662–681. doi: 10.1017/S0305000914000415

Demir-Lira, Ö. E., Asaridou, S. S., Raja Beharelle, A., Holt, A. E., Goldin-Meadow, S., and Small, S. L. (2018). Functional neuroanatomy of gesture-speech integration in children varies with individual differences in gesture processing. *Dev. Sci.* 21:e12648. doi: 10.1111/desc.12648

Dick, A. S., Goldin-Meadow, S., Solodkin, A., and Small, S. L. (2012). Gesture in the developing brain. *Dev. Sci.* 15, 165–180. doi: 10.1111/j.1467-7687.2011.01100.x

Dimeck, P. T., Roy, E. A., and Hall, C. R. (1998). Aging and working memory in gesture imitation. *Brain Cogn.* 37, 124–127.

Dimitrova, N., Özçalışkan, Ş., and Adamson, L. B. (2016). Parents' translations of child gesture facilitate word learning in children with autism, down syndrome and typical development. *J. Autism Dev. Disord.* 46, 221–231. doi: 10.1007/s10803-015-2566-7

Drijvers, L., Vaitonytė, J., and Özyürek, A. (2019). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cogn. Sci.* 43:e12789. doi: 10.1111/cogs.12789

Dror, I. E., and Kosslyn, S. M. (1994). Mental imagery and aging. *Psychol. Aging* 9:90. doi: 10.1037//0882-7974.9.1.90

Eielts, C., Pouw, W., Ouwehand, K., van Gog, T., Zwaan, R. A., and Paas, F. (2018). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychol. Res.* 84, 502–513. doi: 10.1007/s00426-018-1065-9

Evans, J. L., Alibali, M. W., and McNeil, N. M. (2001). Divergence of verbal expression and embodied knowledge: evidence from speech and gesture in children with specific language impairment. *Lang. Cogn. Process.* 16, 309–331. doi: 10.1080/01690960042000049

Feyereisen, P., and Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *J. Nonverbal Behav.* 23, 153–171.

Frick-Horbury, D. (2002). The effects of hand gestures on verbal recall as a function of high-and low-verbal-skill levels. *J. Gen. Psychol.* 129, 137–147. doi: 10.1080/00221300209603134

Galati, A., Weisberg, S. M., Newcombe, N. S., and Avraamides, M. N. (2018). When gestures show us the way: co-thought gestures selectively facilitate navigation and spatial memory. *Spat. Cogn. Comput.* 18, 1–30. doi: 10.1080/13875868.2017.1332064

Gillespie, M., James, A. N., Federmeier, K. D., and Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: evidence from individual differences. *Cognition* 132, 174–180. doi: 10.1016/j.cognition.2014.03.012

Glasser, M. L., Williamson, R. A., and Özçalışkan, Ş. (2018). Do children understand iconic gestures about events as early as iconic gestures about entities? *J. Psycholinguist. Res.* 47, 741–754. doi: 10.1007/s10936-017-9550-7

Glosser, G., Wiley, M. J., and Barnoskir, E. J. (1998). Gestural communication in Alzheimer's disease. *J. Clin. Exp. Neuropsychol.* 20, 1–13. doi: 10.1076/jcen.20.1.1.1484

Göksun, T., Goldin-Meadow, S., Newcombe, N., and Shipley, T. (2013a). Individual differences in mental rotation: what does gesture tell us? *Cogn. Process.* 14, 153–162. doi: 10.1007/s10339-013-0549-1

Göksun, T., Hirsh-Pasek, K., and Golinkoff, R. M. (2010). How do preschoolers express cause in gesture and speech? *Cogn. Dev.* 25, 56–68. doi: 10.1016/j.cogdev.2009.11.001

Göksun, T., Lehet, M., Malykhina, K., and Chatterjee, A. (2013b). Naming and gesturing spatial relations: evidence from focal brain-injured individuals. *Neuropsychologia* 51, 1518–1527. doi: 10.1016/j.neuropsychologia.2013.05.006

Göksun, T., Lehet, M., Malykhina, K., and Chatterjee, A. (2015). Spontaneous gesture and spatial language: evidence from focal brain injury. *Brain Lang.* 150, 1–13. doi: 10.1016/j.bandl.2015.07.012

Goldin-Meadow, S., and Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annu. Rev. Psychol.* 64, 257–283. doi: 10.1146/annurev-psych-113011-143802

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395

Goldin-Meadow, S., and Saltzman, J. (2000). The cultural bounds of maternal accommodation: how Chinese and American mothers communicate with deaf and hearing children. *Psychol. Sci.* 11, 307–314. doi: 10.1111/1467-9280.00261

Graham, J. A., and Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *Eur. J. Soc. Psychol.* 5, 189–195.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish.* Lund: Lund University Press.

Gullberg, M. (2010). Methodological reflections on gesture analysis in second language acquisition and bilingualism research. *Second. Lang. Res.* 26, 75–102. doi: 10.1177/0267658309337639

Gullberg, M., and Holmqvist, K. (1999). Keeping an eye on gestures: visual perception of gestures in face-to-face communication. *Pragmat. Cogn.* 7, 35–63.

Gullberg, M., and Holmqvist, K. (2006). What speakers do and what addressees look at: visual attention to gestures in human interaction live and on video. *Pragmat. Cogn.* 14, 53–82. doi: 10.1075/pc.14.1.05gul

Gullberg, M., and Kita, S. (2009). Attention to speech-accompanying gestures: eye movements and information uptake. *J. Nonverbal Behav.* 33, 251–277. doi: 10.1007/s10919-009-0073-2

Hilverman, C., Cook, S. W., and Duff, M. C. (2018). Hand gestures support word learning in patients with hippocampal amnesia. *Hippocampus* 28, 406–415. doi: 10.1002/hipo.22840

Holle, H., and Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *J. Cogn. Neurosci.* 19, 1175–1192. doi: 10.1162/jocn.2007.19.7.1175

Holler, J., Kendrick, K. H., and Levinson, S. C. (2018). Processing language in face-to-face conversation: questions with gestures get faster responses. *Psychon. Bull. Rev.* 25, 1900–1908. doi: 10.3758/s13423-017-1363-z

Holler, J., Shovelton, H., and Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *J. Nonverbal Behav.* 33, 73–88. doi: 10.1007/s10919-008-0063-9

Holler, J., and Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *J. Lang. Soc. Psychol.* 26, 4–27. doi: 10.1177/0261927X06296428

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol. Bull.* 137, 297–315. doi: 10.1037/a0022128

Hostetter, A. B., and Alibali, M. W. (2007). Raise your hand if you're spatial: relations between verbal and spatial skills and gesture production. *Gesture* 7, 73–95. doi: 10.1075/gest.7.1.05hos

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/pbr.15.3.495

Hostetter, A. B., and Alibali, M. W. (2011). Cognitive skills and gesture-speech redundancy: formulation difficulty or communicative strategy? *Gesture* 11, 40–60. doi: 10.1075/gest.11.1.03hos

Hostetter, A. B., and Alibali, M. W. (2018). Gesture as simulated action: revisiting the framework. *Psychon. Bull. Rev.* 26, 721–752. doi: 10.3758/s13423-018-1548-0

Hostetter, A. B., Murch, S. H., Rothschild, L., and Gillard, C. S. (2018). Does seeing gesture lighten or increase the load? Effects of processing gesture on verbal and visuospatial cognitive load. *Gesture* 17, 268–290. doi: 10.1075/gest.17017.hos

Hostetter, A. B., and Potthoff, A. L. (2012). Effects of personality and social situation on representational gesture production. *Gesture* 12, 62–83. doi: 10.1075/gest.12.1.04hos

Huettig, F., and Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Lang. Cogn. Neurosci.* 31, 80–93. doi: 10.1080/23273798.2015.1047459

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *J. Comput. Assist. Learn.* 22, 392–404. doi: 10.1111/j.1365-2729.2006.00180.x

Iverson, J. M., and Braddock, B. A. (2011). Gesture and motor skill in relation to language in children with language impairment. *J. Speech Lang. Hear. Res.* 54, 72–86. doi: 10.1044/1092-4388(2010/08-0197)

Iverson, J. M., Capirci, O., Longobardi, E., and Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cogn. Dev.* 14, 57–75.

Iverson, J. M., Capirci, O., Volterra, V., and Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: early communication in Italian vs. American children. *First Lang.* 28, 164–181. doi: 10.1177/0142723707087736

Iverson, J. M., Longobardi, E., Spampinato, K., and Caselli, M. C. (2006). Gesture and speech in maternal input to children with Down's syndrome. *Int. J. Lang. Commun. Disord.* 41, 235–251. doi: 10.1080/13682820500312151

Jorm, A. F., Korten, A. E., and Henderson, A. S. (1987). The prevalence of dementia: a quantitative integration of the literature. *Acta Psychiatr. Scand.* 76, 465–479. doi: 10.1111/j.1600-0447.1987.tb02906.x

Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122–149. doi: 10.1037/0033-295x.99.1.122

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educ. Psychol. Rev.* 19, 509–539. doi: 10.1007/s10648-007-9054-3

Kane, M. J., and Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671. doi: 10.3758/bf03196323

Karadöller, D. Z., Ünal, E., Sumer, B., Göksun, T., Özer, D., and Ozyurek, A. (2019). "Children but not adults use both speech and gesture to produce informative expressions of left-right relations" in *the 44th Annual Boston University Conference on Language Development (BUCLD 44)*; October 7-10, 2019.

Kartalkanat, H., and Göksun, T. (2020). The effects of observing different gestures during storytelling on the recall of path and event information in 5-year-olds and adults. *J. Exp. Child Psychol.* 189:104725. doi: 10.1016/j.jecp.2019.104725

Kelly, S. D., and Goldsmith, L. H. (2004). Gesture and right hemisphere involvement in evaluating lecture material. *Gesture* 4, 25–42. doi: 10.1075/gest.4.1.03kel

Kelly, S. D., Hirata, Y., Manansala, M., and Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Front. Psychol.* 5:673. doi: 10.3389/fpsyg.2014.00673

Kelly, S. D., Kravitz, C., and Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* 89, 253–260. doi: 10.1016/S0093-934X(03)00335-3

Kelly, S. D., Manning, S. M., and Rodak, S. (2008). Gesture gives a hand to language and learning: perspectives from cognitive neuroscience, developmental psychology, and education. *Lang Ling Compass* 2, 569–588. doi: 10.1111/j.1749-818X.2008.00067.x

Kelly, S. D., Özyürek, A., and Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kiat, J. E., and Belli, R. F. (2018). The role of individual differences in visual\verbal information processing preferences in visual\verbal source monitoring. *J. Cogn. Psychol.* 30, 701–709. doi: 10.1080/20445911.2018.1509865

Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends Cogn. Sci.* 22, 154–169. doi: 10.1016/j.tics.2017.11.006

Kim, Z. H., and Lausberg, H. (2018). Koreans and germans: cultural differences in hand movement behaviour and gestural repertoire. *J. Intercult. Commun. Res.* 47, 439–453. doi: 10.1080/17475759.2018.1475296

Kirby, J., Moore, P., and Shofield, N. (1988). Verbal and visual learning styles. *Contemp. Educ. Psychol.* 13, 169–184.

Kirk, E., Pine, K. J., and Ryder, N. (2011). I hear what you say but I see what you mean: the role of gestures in children's pragmatic comprehension. *Lang. Cogn. Process.* 26, 149–170. doi: 10.1080/01690961003752348

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Comput. Educ.* 106, 166–171. doi: 10.1016/j.compedu.2016.12.006

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Lang. Cogn. Process.* 24, 145–167. doi: 10.1080/01690960802586188

Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059

Kita, S., and Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Lang. Cogn. Process.* 24, 761–775. doi: 10.1080/01690960802327971

Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3

Klooster, N. B., Cook, S. W., Uc, E. Y., and Duff, M. C. (2015). Gestures make memories, but what kind? Patients with impaired procedural memory display disruptions in gesture production and comprehension. *Front. Hum. Neurosci.* 8:1054. doi: 10.3389/fnhum.2014.01054

Koć-Januchta, M., Höffler, T., Thoma, G. B., Prechtl, H., and Leutner, D. (2017). Visualizers versus verbalizers: effects of cognitive style on learning with texts and pictures—an eye-tracking study. *Comput. Hum. Behav.* 68, 170–179. doi: 10.1016/j.chb.2016.11.028

Kozhevnikov, M., Hegarty, M., and Mayer, R. E. (2002). Revising the visualizer/verbalizer dimension: evidence for two types of visualizers. *Cogn. Instr.* 20, 47–77. doi: 10.1207/S1532690XCI2001_3

Krauss, R. M., Chen, Y., and Gottesman, R. F. (2000). "Lexical gestures and lexical access: a process model" in *Language and gesture*. ed. D. McNeill (Cambridge, UK: Cambridge University Press).

Krauss, R., and Hadar, U. (1999). "The role of speech-related arm/hand gesture in word retrieval" in *Gesture, speech, and sign*. eds. L. S. Messing and R. Campbell (Oxford: Oxford University Press).

Lausberg, H., and Kita, S. (2003). The content of the message influences the hand choice in co-speech gestures and in gesturing without speaking. *Brain Lang.* 86, 57–69. doi: 10.1016/s0093-934x(02)00534-5

LeBarton, E. S., and Iverson, J. M. (2017). "Gesture's role in learning interactions" in *Why gesture? How the hands function in speaking, thinking and communicating*. eds. R. B. Church, M. W. Alibali and S. D. Kelly (Amsterdam: John Benjamins Publishing), 331–351.

Liszkowski, U., Carpenter, M., and Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition* 108, 732–739. doi: 10.1016/j.cognition.2008.06.013

Macoun, A., and Sweller, N. (2016). Listening and watching: the effects of observing gesture on preschoolers' narrative comprehension. *Cogn. Dev.* 40, 68–81. doi: 10.1016/j.cogdev.2016.08.005

Mainela-Arnold, E., Alibali, M. W., Hostetter, A. B., and Evans, J. L. (2014). Gesture-speech integration in children with specific language impairment. *Int. J. Lang. Commun. Disord.* 49, 761–770. doi: 10.1111/1460-6984.12115

Mainela-Arnold, E., Alibali, M. W., Ryan, K., and Evans, J. L. (2011). Knowledge of mathematical equivalence in children with specific language impairment: insights from gesture and speech. *Lang. Speech Hear. Serv. Sch.* 42, 18–30. doi: 10.1044/0161-1461(2010/09-0070)

Marstaller, L., and Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychon. Bull. Rev.* 20, 496–500. doi: 10.3758/s13423-012-0365-0

Mastrogiuseppe, M., Capirci, O., Cuva, S., and Venuti, P. (2015). Gestural communication in children with autism spectrum disorders during mother-child interaction. *Autism* 19, 469–481. doi: 10.1177/1362361314528390

Mayberry, R. I., and Nicoladis, E. (2000). Gesture reflects language development: evidence from bilingual children. *Curr. Dir. Psychol. Sci.* 9, 192–196. doi: 10.1111/1467-8721.00092

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.

McNeill, D., and Duncan, S. D. (2000). "Growth points in thinking-for-speaking" in *Language and gesture*. ed. D. McNeill (Cambridge: Cambridge University Press), 141–161.

Melinger, A., and Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture* 4, 119–141. doi: 10.1075/gest.4.2.02mel

Mendelson, A. L., and Thorson, E. (2004). How verbalizers and visualizers process the newspaper environment. *J. Commun.* 54, 474–491. doi: 10.1111/j.1460-2466.2004.tb02640.x

Meneghetti, C., Labate, E., Grassano, M., Ronconi, L., and Pazzaglia, F. (2014). The role of visuospatial and verbal abilities, styles and strategies in predicting visuospatial description accuracy. *Learn. Individ. Differ.* 36, 117–123. doi: 10.1016/j.lindif.2014.10.019

Montepare, J., Koff, E., Zaitchik, D., and Albert, M. L. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *J. Nonverbal Behav.* 23, 133–152.

Morsella, E., and Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *Am. J. Psychol.* 117, 411–424. doi: 10.2307/4149008

Mundy, P., Sigman, M., and Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *J. Autism Dev. Disord.* 20, 115–128. doi: 10.1007/BF02206861

Nagels, A., Kircher, T., Steines, M., Grosvald, M., and Straube, B. (2015). A brief self-rating scale for the assessment of individual differences in gesture perception and production. *Learn. Individ. Differ.* 39, 73–80. doi: 10.1016/j.lindif.2015.03.008

Nagpal, J., Nicoladis, E., and Marentette, P. (2011). Predicting individual differences in L2 speakers' gestures. *Int. J. Biling.* 15, 205–214. doi: 10.1177/1367006910381195

Newcombe, N. S., Uttal, D. H., and Sauter, M. (2013). "Spatial development" in *Oxford library of psychology. The Oxford handbook of developmental psychology (Vol. 1): Body and mind*. ed. P. D. Zelazo (New York, NY: Oxford University Press), 564–590.

Nicoladis, E., Pika, S., and Marentette, P. (2009). Do French–English bilingual children gesture more than monolingual children? *J. Psycholinguist. Res.* 38, 573–585. doi: 10.1007/s10936-009-9121-7

Novack, M. A., and Goldin-Meadow, S. (2017). Gesture as representational action: a paper about function. *Psychon. Bull. Rev.* 24, 652–665. doi: 10.3758/s13423-016-1145-z

Obermeier, C., Dolk, T., and Gunter, T. C. (2012). The benefit of gestures during communication: evidence from hearing and hearing-impaired individuals. *Cortex* 48, 857–870. doi: 10.1016/j.cortex.2011.02.007

Özçalışkan, Ş., Adamson, L. B., and Dimitrova, N. (2016). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism* 20, 754–763. doi: 10.1177/1362361315605921

Özçalışkan, Ş., Adamson, L. B., Dimitrova, N., and Baumann, S. (2017). Early gesture provides a helping hand to spoken vocabulary development for children with autism, Down syndrome, and typical development. *J. Cogn. Dev.* 18, 325–337. doi: 10.1080/15248372.2017.1329735

Özçalışkan, Ş., Adamson, L. B., Dimitrova, N., and Baumann, S. (2018). Do parents model gestures differently when children's gestures differ? *J. Autism Dev. Disord.* 48, 1492–1507. doi: 10.1007/s10803-017-3411-y

Özçalışkan, Ş., and Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition* 96, 101–113. doi: 10.1016/j.cognition.2005.01.001

Özçalışkan, Ş., and Goldin-Meadow, S. (2010). Sex differences in language first appear in gesture. *Dev. Sci.* 13, 752–760. doi: 10.1111/j.1467-7687.2009.00933.x

Özçalışkan, Ş., Levine, S. C., and Goldin-Meadow, S. (2013). Gesturing with an injured brain: how gesture helps children with early brain injury learn linguistic constructions. *J. Child Lang.* 40:69. doi: 10.1017/S0305000912000220

Özer, D., and Göksun, T. (2020). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Lang. Cogn. Neurosci.* 35, 896–914. doi: 10.1080/23273798.2019.1703016

Özer, D., Göksun, T., and Chatterjee, A. (2019). Differential roles of gestures on spatial language in neurotypical elderly adults and individuals with focal brain injury. *Cogn. Neuropsychol.* 36, 282–299. doi: 10.1080/02643294.2019.1618255

Özer, D., Tansan, M., Özer, E. E., Malykhina, K., Chatterjee, A., and Göksun, T. (2017). "The effects of gesture restriction on spatial language in young and elderly adults" in *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. eds. G. Gunzelmann, A. Howes, T. Tenbrink and E. Davelaar. July 26-29, 2017; (Austin, TX: Cognitive Science Society), 1471–1476.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philos. Trans. R. Soc. B.: Biol. Sci.* 369:20130296. doi: 10.1098/rstb.2013.0296

Perry, M., Church, R. B., and Goldin-Meadow, S. (1992). Is gesture-speech mismatch a general index of transitional knowledge? *Cogn. Dev.* 7, 109–122.

Pika, S., Nicoladis, E., and Marentette, P. F. (2006). A cross-cultural study on the use of gestures: evidence for cross-linguistic transfer? *Biling.: Lang. Cogn.* 9, 319–327. doi: 10.1017/S1366728906002665

Ping, R., and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cogn. Sci.* 34, 602–619. doi: 10.1111/j.1551-6709.2010.01102.x

Post, L. S., Van Gog, T., Paas, F., and Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Comput. Hum. Behav.* 29, 1450–1455. doi: 10.1016/j.chb.2013.01.005

Pouw, W. T., De Nooijer, J. A., Van Gog, T., Zwaan, R. A., and Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Front. Psychol.* 5:359. doi: 10.3389/fpsyg.2014.00359

Pouw, W. T., Mavilidi, M. F., van Gog, T., and Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cogn. Process.* 17, 269–277. doi: 10.1007/s10339-016-0757-6

Priesters, M. A., and Mittelberg, I. (2013). "Individual differences in speakers' gesture spaces: multi-angle views from a motion-capture study" in *Proceedings of the Tilburg Gesture Research Meeting (TiGeR)*; June 19-21, 2013; 19–21.

Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). Gesture, speech, and lexical access: the role of lexical movements in speech production. *Psychol. Sci.* 7, 226–231.

Richmond, V. P., McCroskey, J. C., and Johnson, A. D. (2003). Development of the nonverbal immediacy scale (NIS): measures of self-and other-perceived nonverbal immediacy. *Commun. Q.* 51, 504–517. doi: 10.1080/01463370309370170

Riding, R., Burton, D., Rees, G., and Sharratt, M. (1995). Cognitive style and personality in 12-year-old children. *Br. J. Educ. Psychol.* 65, 113–124. doi: 10.1111/j.2044-8279.1995.tb01135.x

Roth, W. M. (2001). Gestures: their role in teaching and learning. *Rev. Educ. Res.* 71, 365–392. doi: 10.3102/00346543071003365

Rousseaux, M., Rénier, J., Anicet, L., Pasquier, F., and Mackowiak-Cordoliani, M. A. (2012). Gesture comprehension, knowledge and production in Alzheimer's disease. *Eur. J. Neurol.* 19, 1037–1044. doi: 10.1111/j.1468-1331.2012.03674.x

Rowe, M. L., and Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Dev. Sci.* 12, 182–187. doi: 10.1111/j.1467-7687.2008.00764.x

Rowe, M. L., Özçalışkan, Ş., and Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Lang.* 28, 182–199. doi: 10.1177/0142723707088310

Rozga, A., Hutman, T., Young, G. S., Rogers, S. J., Ozonoff, S., Dapretto, M., et al. (2011). Behavioral profiles of affected and unaffected siblings of children with autism: contribution of measures of mother-infant interaction and nonverbal communication. *J. Autism Dev. Disord.* 41, 287–301. doi: 10.1007/s10803-010-1051-6

Rueckert, L., Church, R. B., Avila, A., and Trejo, T. (2017). Gesture enhances learning of a complex statistical concept. *Cogn. Res. Princ. Implic.* 2:2. doi: 10.1186/s41235-016-0036-1

Sassenberg, U., Foth, M., Wartenburger, I., and van der Meer, E. (2011). Show your hands—are you really clever? Reasoning, gesture production, and intelligence. *Linguist.* 49, 105–134. doi: 10.1515/ling.2011.003

Sauer, E., Levine, S. C., and Goldin-Meadow, S. (2010). Early gesture predicts language delay in children with pre-or perinatal brain lesions. *Child Dev.* 81, 528–539. doi: 10.1111/j.1467-8624.2009.01413.x

Schmalenbach, S. B., Billino, J., Kircher, T., van Kemenade, B. M., and Straube, B. (2017). Links between gestures and multisensory processing: individual differences suggest a compensation mechanism. *Front. Psychol.* 8:1828. doi: 10.3389/fpsyg.2017.01828

Schubotz, L., Özyürek, A., and Holler, J. (2019). Age-related differences in multimodal recipient design: younger, but not older adults, adapt speech and co-speech gestures to common ground. *Lang. Cogn. Neurosci.* 34, 254–271. doi: 10.1080/23273798.2018.1527377

Sekine, K., and Kita, S. (2015). Development of multimodal discourse comprehension: cohesive use of space by gestures. *Lang. Cogn. Neurosci.* 30, 1245–1258. doi: 10.1080/23273798.2015.1053814

Sekine, K., Schoechl, C., Mulder, K., Holler, J., Kelly, S., Furman, R., et al. (in press). Evidence for children's online integration of simultaneous information from speech and iconic gestures: an ERP study. *Lang. Cogn. Neurosci.* 1–12.

Sekine, K., Sowden, H., and Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cogn. Sci.* 39, 1855–1880. doi: 10.1111/cogs.12221

Sheehan, E. A., Namy, L. L., and Mills, D. L. (2007). Developmental changes in neural activity to familiar words and gestures. *Brain Lang.* 101, 246–259. doi: 10.1016/j.bandl.2006.11.008

Ska, B., and Croisile, B. (1998). Effects of normal aging on the recognition of gestures. *Brain Cogn.* 37, 136–138.

Smithson, L., and Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Biling.: Lang. Cogn.* 16, 934–944. doi: 10.1017/S1366728913000175

Smithson, L., Nicoladis, E., and Marentette, P. (2011). Bilingual children's gesture use. *Gesture* 11, 330–347. doi: 10.1075/gest.11.3.04smi

So, W. C., Kita, S., and Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cogn. Sci.* 33, 115–125. doi: 10.1111/j.1551-6709.2008.01006.x

So, W. C., Shum, P. L. C., and Wong, M. K. Y. (2015). Gesture is more effective than spatial language in encoding spatial information. *Q. J. Exp. Psychol.* 68, 2384–2401. doi: 10.1080/17470218.2015.1015431

Sowden, H., Clegg, J., and Perkins, M. (2013). The development of co-speech gesture in the communication of children with autism spectrum disorders. *Clin. Linguist. Phon.* 27, 922–939. doi: 10.3109/02699206.2013.818715

Stanfield, C., Williamson, R., and Özçalışkan, Ş. E. Y. D. A. (2014). How early do children understand gesture-speech combinations with iconic gestures? *J. Child Lang.* 41, 462–471. doi: 10.1017/S0305000913000019

Stefanini, S., Caselli, M. C., and Volterra, V. (2007). Spoken and gestural production in a naming task by young children with down syndrome. *Brain Lang.* 101, 208–221. doi: 10.1016/j.bandl.2007.01.005

Stone, W. L., Ousley, O. Y., Yoder, P. J., Hogan, K. L., and Hepburn, S. L. (1997). Nonverbal communication in two-and three-year-old children with autism. *J. Autism Dev. Disord.* 27, 677–696. doi: 10.1023/a:1025854816091

Streeck, J. (2009). Depicting gestures: examples of the analysis of embodied communication in the arts of the west. *Gesture* 9, 1–34. doi: 10.1075/gest.9.1.01str

Talbott, M. R., Nelson, C. A., and Tager-Flusberg, H. (2015). Maternal gesture use and language development in infant siblings of children with autism spectrum disorder. *J. Autism Dev. Disord.* 45, 4–14. doi: 10.1007/s10803-013-1820-0

Tamis-LeMonda, C. S., Song, L., Leavell, A. S., Kahana-Kalman, R., and Yoshikawa, H. (2012). Ethnic differences in mother-infant language and gestural communications are associated with specific skills in infants. *Dev. Sci.* 15, 384–397. doi: 10.1111/j.1467-7687.2012.01136.x

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: a comparison between young and older adults. *Psychol. Aging* 10, 215–228. doi: 10.1037//0882-7974.10.2.215

Thompson, L. A., and Guzman, F. A. (1999). Some limits on encoding visible speech and gestures using a dichotic shadowing task. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 54, P347–P349. doi: 10.1093/geronb/54b.6.p347

Trafton, J. G., Trickett, S. B., Stitzlein, C. A., Saner, L., Schunn, C. D., and Kirschenbaum, S. S. (2006). The relationship between spatial transformations and iconic gestures. *Spat. Cogn. Comput.* 6, 1–29. doi: 10.1207/s15427633scc0601_1

Trujillo, J. P., Simanova, I., Bekkering, H., and Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: a Kinect study. *Cognition* 180, 38–51. doi: 10.1016/j.cognition.2018.04.003

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *Am. Psychol.* 30:128,

van Wermeskerken, M., Fijan, N., Eielts, C., and Pouw, W. T. (2016). Observation of depictive versus tracing gestures selectively aids verbal versus visual-spatial learning in primary school children. *Appl. Cogn. Psychol.* 30, 806–814. doi: 10.1002/acp.3256

Vanetti, E. J., and Allen, G. L. (1988). Communicating environmental knowledge: the impact of verbal and spatial abilities on the production and comprehension of route directions. *Environ. Behav.* 20, 667–682.

Vogel, E. K., and Awh, E. (2008). How to exploit diversity for scientific gain: using individual differences to constrain cognitive theory. *Curr. Dir. Psychol. Sci.* 17, 171–176. doi: 10.1111/j.1467-8721.2008.00569.x

Vogt, S., and Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *J. Child Lang.* 44, 1458–1484. doi: 10.1017/S0305000916000647

Wakefield, E. M., James, T. W., and James, K. H. (2013). Neural correlates of gesture processing across human development. *Cogn. Neuropsychol.* 30, 58–76. doi: 10.1080/02643294.2013.794777

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., and Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci.* 21:e12664. doi: 10.1111/desc.12664

Wartenburger, I., Kühn, E., Sassenberg, U., Foth, M., Franz, E. A., and van der Meer, E. (2010). On the relationship between fluid intelligence, gesture production, and brain structure. *Intelligence* 38, 193–201. doi: 10.1016/j.intell.2009.11.001

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., and Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: a retrospective home video study. *Am. J. Speech Lang. Pathol.* 22, 25–39. doi: 10.1044/1058-0360(2012/11-0145)

Wermelinger, S., Gampe, A., Helbling, N., and Daum, M. M. (2020). Do you understand what I want to tell you? Early sensitivity in bilinguals' iconic gesture perception and production. *Dev. Sci.* 23:e12943. doi: 10.1111/desc.12943

Wesp, R., Hesse, J., Keutmann, D., and Wheaton, K. (2001). Gestures maintain spatial imagery. *Am. J. Psychol.* 114, 591–600. doi: 10.2307/1423612

Willems, R. M., Özyürek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage* 47, 1992–2004. doi: 10.1016/j.neuroimage.2009.05.066

Wu, Y. C., and Coulson, S. (2005). Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42, 654–667. doi: 10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., and Coulson, S. (2014a). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychol.* 153, 39–50. doi: 10.1016/j.actpsy.2014.09.002

Wu, Y. C., and Coulson, S. (2014b). A psychometric measure of working memory capacity for configured body movement. *PLoS One* 9:e84834. doi: 10.1371/journal.pone.0084834

Yeo, L. M., and Tzeng, Y. T. (2019). Tracing effect in the worked examples-based learning: an exploration of individual differences in working memory capacity. *Eurasia J. Math. Sci. Technol. Educ.* 15:em1760. doi: 10.29333/ejmste/105482

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 reviewer declared a shared collaboration in a society with one of the authors TG at time of review.

frontiers in Psychology

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/fpsyg.2020.575032

Using Gesture to Facilitate L2 Phoneme Acquisition: The Importance of Gesture and Phoneme Complexity

Marieke Hoetjes and Lieke van Maastricht*

Centre for Language Studies, Radboud University, Nijmegen, Netherlands

Most language learners have difficulties acquiring the phonemes of a second language (L2). Unfortunately, they are often judged on their L2 pronunciation, and segmental inaccuracies contribute to miscommunication. Therefore, we aim to determine how to facilitate phoneme acquisition. Given the close relationship between speech and co-speech gesture, previous work unsurprisingly reports that gestures can benefit language acquisition, e.g., in (L2) word learning. However, gesture studies on L2 phoneme acquisition present contradictory results, implying that both specific properties of gestures and phonemes used in training, and their combination, may be relevant. We investigated the effect of phoneme and gesture complexity on L2 phoneme acquisition. In a production study, Dutch natives received instruction on the pronunciation of two Spanish phonemes, /u/ and /θ/. Both are typically difficult to produce for Dutch natives because their orthographic representation differs between both languages. Moreover, /θ/ is considered more complex than /u/, since the Dutch phoneme inventory contains /u/ but not /θ/. The instruction participants received contained Spanish examples presented either via audio-only, audio-visually without gesture, audio-visually with a simple, pointing gesture, or audio-visually with a more complex, iconic gesture representing the relevant speech articulator(s). Preceding and following training, participants read aloud Spanish sentences containing the target phonemes. In a perception study, Spanish natives rated the target words from the production study on accentedness and comprehensibility. Our results show that combining gesture and speech in L2 phoneme training can lead to significant improvement in L2 phoneme production, but both gesture and phoneme complexity affect successful learning: Significant learning only occurred for the less complex phoneme /u/ after seeing the more complex iconic gesture, whereas for the more complex phoneme /θ/, seeing the more complex gesture actually hindered acquisition. The perception results confirm the production findings and show that items containing /θ/ produced after receiving training with a less complex pointing gesture are considered less foreign-accented and more easily comprehensible as compared to the same items after audio-only training. This shows that gesture can facilitate task performance in L2 phonology acquisition, yet complexity affects whether certain gestures work better for certain phonemes than others.

Keywords: second language acquisition, phonemes, audiovisual, deictic gesture, iconic gesture, accentedness, comprehensibility

OPEN ACCESS

Edited by:
Kazuki Sekine,
Waseda University, Japan

Reviewed by:
Gaelle Ferre,
University of Poitiers, France
Peng Li,
Pompeu Fabra University, Spain

****Correspondence:***
Marieke Hoetjes
m.hoetjes@let.ru.nl

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

Received: *22 June 2020*
Accepted: *26 October 2020*
Published: *23 November 2020*

Citation:
Hoetjes M and van Maastricht L (2020) Using Gesture to Facilitate L2 Phoneme Acquisition: The Importance of Gesture and Phoneme Complexity. Front. Psychol. 11:575032. doi: 10.3389/fpsyg.2020.575032

Preliminary versions of parts of this paper were presented at the International Congress of Phonetic Sciences in August 2019 in Melbourne, Australia (Van Maastricht et al., 2019), at the 29th conference of the European Second Language Association in August 2019 in Lund, Sweden (Hoetjes et al., 2019b), and at the Gesture and Speech in Interaction conference in September 2019 in Paderborn, Germany (Hoetjes et al., 2019a). The current paper includes a more detailed theoretical background, description of the experimental methods, and discussion of the findings, as well as more advanced statistical analyses over the complete data set in the case of Study I and analyses over a new data set in the case of Study II.

INTRODUCTION

Human communication is multimodal: When people communicate face-to-face, they do not only use speech but also various non-verbal communicative cues, such as facial expressions and hand gestures. In this study, we focus on one of these aspects of non-verbal communication, namely co-speech hand gestures, within the context of foreign language learning. There is general agreement in the literature that speech and co-speech gestures are closely related and that they are integrated in various ways (McNeill, 1992; Kendon, 2004; Wagner et al., 2014). This is apparent, for example, by the fact that there is a close temporal and semantic coordination between speech and gesture. This means that roughly speaking, speech and gesture tend to express the same thing at the same time (see, Gullberg, 2006, for an overview). Moreover, the integration between speech and gesture is reflected in the parallel development of the two modalities: For instance, in first language (L1) acquisition, it has been shown that gestures play a facilitating role in vocabulary learning in children, with gesture production predicting their subsequent lexical and syntactic development (e.g., Goldin-Meadow, 2005). Both modalities have also been shown to break down in a parallel way, for example during disfluencies (e.g., Seyfeddinipur, 2006; Graziano and Gullberg, 2018) or as a result of aphasia (Van Nispen et al., 2016). In short, the relationship between speech and gesture plays a crucial role in our communicative processes. Given this close relationship between speech and gesture in communication, the possible benefit of gesture in learning contexts has been a topic of research in different scientific fields, one of which is second language (L2) acquisition. While gesture is often intuitively used by teachers in classrooms (cf. Smotrova, 2017), very little is known about the specifics of the interplay between both modalities in a learning context. Hence, in the current study, we compare the use of different types of gestures in the context of L2 phoneme acquisition to determine in which way gesture and phoneme complexity in L2 training affect the phoneme productions of Dutch learners of Spanish (Study I) and the perceptions of Spanish natives with respect to these non-native productions (Study II). Before turning to the specifics of our research, we first review the relevant literature.

Multimodality in Learning Contexts

Gesture can play a facilitative role in various kinds of learning situations. For example, previous work has shown that students take teachers' gestures into account and that teachers can thus use gesture to help students learn mathematical concepts (e.g., Goldin-Meadow et al., 1999; Yeo et al., 2018). Focusing on L2 learning, various studies have shown that gestures can play a facilitative role in the acquisition of L2 vocabulary, both by children and adults. Tellier (2008), for example, had 5-year old French children learn English words associated with either a picture or a gesture and found that the gesture group did better than the picture group. For adults, Kelly et al. (2009) likewise found that when novel Japanese words were presented to native speakers of English, they were better at learning these words when they were presented with hand gestures, as compared to without hand gestures. In these studies, iconic gestures were used, which have a clear semantic relationship to the lexical items they accompany. The conclusion we can draw from these findings is that presenting semantic information in several modalities strengthens learners' memory of the words' semantic meaning (e.g., Tellier, 2008; Kelly et al., 2009; Macedonia et al., 2011).

Apart from vocabulary acquisition, it is important for L2 learners to also learn how to correctly pronounce the sounds of their target language. On the one hand, phoneme acquisition is one of the aspects of L2 acquisition learners generally find most difficult (see, e.g., collected papers in Bohn and Munro, 2007), while on the other hand, an atypical pronunciation is an aspect of speech that is very salient to native listeners (see Derwing and Munro, 2009 and the references therein), even if it doesn't necessarily affect their perceived ease of comprehensibility or actual processing of the L2 speech (Munro and Derwing, 1999; Van Maastricht et al., 2016). Moreover, pronunciation is often one of the aspects of the L2 that learners are eager to acquire since most of them aim to sound as native-like as possible in the L2 (Timmis, 2002; Derwing, 2003). A native-like pronunciation is especially important given that a clear non-native pronunciation has been shown to negatively affect the way speakers are perceived (Lev-Ari and Keysar, 2010) and segmental inaccuracies contribute to miscommunication (Caspers and Horłoza, 2012).

Given the tight relationship between speech and gesture and the fact that gestures can facilitate L1, and even L2, development, it is not such a strange idea that gesture may also play a role in L2 phoneme acquisition. Anecdotally, L2 teachers report to regularly use gestures in the classroom when teaching different aspects of L2 phonology but there are also empirical reasons to assume that gestures could play a facilitative role in L2 phoneme acquisition even though, to date, most research on multimodal L2 phonology acquisition has not focused on gestures. For instance, Hazan et al. (2005) have shown that multimodal training on English phoneme contrasts, in this case through the auditory modality only as compared to through the audiovisual modality, generally benefitted the production and perception of L2 phonemes by Japanese learners of English. Hardison (2003) reports similar results with Japanese and Korean intermediate-level learners of English and found that improvement in phoneme perception also led to improved phoneme production, which she attributes to the fact that the audiovisual training leaves multiple memory traces, while the auditory training only left one.

Using a form of multimodal training that is similar to a gesture, Zhang et al. (2020) studied the facilitative effect of hand-clapping on L2 pronunciation. They showed that French words produced by Chinese adolescents were rated as marginally more nativelike after they had seen and reproduced training videos in which the speaker clapped to visualize the rhythmic structure of the French words as compared to seeing a speaker that did not move her hands and not moving their own hands. They also found a significant effect of training condition on final syllable duration, reflecting the final stress placement that is typical of French, with longer final syllable lengths for items produced after the clapping condition. Like hand-clapping, gestures are not only visual but also consist of movements. Hence, these previous findings would suggest that using gesture in language training, as opposed to using only auditory input or visual input without movements, could facilitate L2 phoneme acquisition. Indeed, some previous studies have been conducted specifically on the role of gestures in the acquisition of L2 tonal and phonemic contrasts. However, the results of these studies are inconclusive.

Gesture and L2 Phonology

On the one hand, there is previous work suggesting that gestures can indeed play a role in the acquisition of certain aspects of L2 phonology, such as the perception of L2 tones and intonation contours. Kelly et al. (2017) conducted a study in which native speakers of English listened to different types of Japanese phonemic contrasts. The speech sounds contrasted concerning their vowel length or their sentence-final intonation. Participants were presented with training on the relevant phonemic differences, followed by videos showing either speech without gestures, speech with congruent metaphoric gestures visualizing the contrast, where the gestures' meaning was in line with the phonemic meaning (short vs. long vowel, or rising vs. falling intonation), or speech with incongruent gestures (e.g., a short vowel with a long gesture). After each video, participants had to indicate whether they perceived the audio to contain a long vs. short vowel, or rising vs. falling intonation. Although results were not clear-cut for the vowel length contrasts, congruent gestures did help to correctly perceive intonational contrasts, as compared to incongruent gesture or no gesture conditions. In a similar vein, work by Hannah et al. (2017) on Mandarin tones used speech-accompanying congruent and incongruent metaphoric gestures and found that perceivers often relied on the visual cues they received, which in the case of incongruence between speech and gesture resulted in participants incorrectly perceiving what they had heard. Gluhareva and Prieto (2017) did not use metaphoric gestures but beat gestures, and showed that viewing beat gestures during discourse prompts improved L2 pronunciation, as measured by accentedness ratings by English natives of short stories produced by Catalan learners of English. Moreover, recent work by Li et al. (2020) focused on the L2 acquisition of Japanese vowel-length contrasts and although they found that gesture (versus no gesture) did not improve L2 vowel length perception, gesture did facilitate correct L2 vowel length production.

On the other hand, there has been work suggesting that gestures do not play a facilitative role in the acquisition of some aspects of L2 phonology, such as the perception of phonemic vowel length distinctions in Kelly et al. (2017), where viewing gestures did not facilitate the perception of phonemic vowel length distinctions. Several other studies also did not report positive effects of gesture on L2 phoneme perception. For instance, in work by Kelly et al. (2014) and by Hirata et al. (2014), the L2 acquisition of phonemic vowel length contrasts was investigated by letting English naïve learners of Japanese observe or also produce gestures related to the syllable or the mora structure of the target word. In an auditory identification task, no differences between the training conditions were found. The authors suggest that this could mean that gestures are not suited for learning phonetic distinctions[1]. Earlier work by Kelly and Lee (2012) expounds this point of view somewhat by stating that gesture may help in acquiring phonetically easy phonemic contrasts, but hinders the acquisition of phonetically hard contrasts because iconic gestures could add too much semantic content to the spoken input, which complicates the acquisition of new phonemes since the learner is simultaneously paying attention to the novel sounds and the contents of the gesture. Hence, they suggest that "it is possible that gesture facilitates local processing of speech sounds only for familiar phonemes in one's native language" (p. 804), which is a relevant factor in the present study.

This contrast between gestures playing a facilitative role in certain contexts but hindering L2 acquisition in others has, in some cases, even been shown within studies. As discussed above, Kelly et al. (2017), for example, showed that similar metaphoric gestures helped for perceiving non-native intonation contours, but did not help in perceiving vowel length differences. Likewise, Morett and Chang (2015) studied the acquisition of L2 Mandarin lexical tone perception by English learners and found that gestures that visualize the target pitch contour helped acquisition, while gestures referring to the semantic meaning of the word hindered correct tone identification. Clearly, the role of gestures in the L2 acquisition of phonemes is not straight-forward. As prior studies used varying research methods and focused on different aspects of L2 phonology, it remains unclear whether the contradictory findings within the field of L2 phonology acquisition are due to methodological discrepancies or to the fact that the specific properties of the gestures used in training, as well as the properties of the phonetic feature to be acquired, contribute to the effectiveness of the use of gesture in L2 pronunciation training. It has been suggested (Kelly et al., 2014) that using gestures for complex L2 input, for example, because the learner has a low proficiency or because the contrast in question is hard to acquire, may hinder rather than help acquisition. In those cases, the processing resources needed for the interpretation of the speech might be prioritized to those needed to process the gesture. This would be in contrast with easy L2 acquisition contexts, where gestures that may play a beneficial role can be processed alongside speech. In any case, the lack of agreement

[1] Alternatively, the lack of effect in the perception task may also be due to a ceiling effect (cf. Hayes-Harb and Masuda, 2008; Li et al., 2020).

between the different studies in this domain means that it is hard to draw clear conclusions, and indeed, Kelly et al. (2017, p. 1) suggest that "gestures help with some –but not all- novel speech sounds in a foreign language."

The Present Study

What most previous studies on L2 phoneme acquisition have in common is that they generally focus on learners' *perception* skills, that is, whether certain types of language training result in learners being able to recognize or distinguish between different phonemes. In most cases, we do not yet know to what extent these results can be extended to learners' *production* of L2 phonemes. In other words, can a certain type of training result in L2 learners' improved ability to pronounce the phonemes in the L2? Hence, one of the goals of this study is to focus on L2 phoneme production. Also, one potential reason for the diverging findings in previous work is that the effect of gestures in L2 phoneme training on L2 phoneme perception has been investigated using various types of gestures and hand movements, but without directly comparing them. Studies have, for example, looked at the use of beats (Gluhareva and Prieto, 2017), which are simple rhythmical gestures, but also at, arguably more complex, metaphoric gestures (Kelly et al., 2014, 2017), which are like iconic gestures in the sense that they show a clear semantic relationship between the movement and the content of speech, but are produced during abstract speech. We are unaware of previous work incorporating deictic (i.e., pointing) gestures in L2 acquisition or of work on L2 phoneme acquisition comparing the effect of different types of gestures. These differences between studies make it hard to draw clear conclusions about the educational value of different types of gestures. Differences in the speech-gesture relationship between types of gestures mean that their potential role in L2 acquisition is not self-evident. Hence, another goal of this study is to compare different *types of gestures* and the role they may play in the acquisition of L2 phoneme production.

In the current study, we thus aim to investigate whether different types of gestures can facilitate L2 learners' productions of two different L2 phonemes, which vary in complexity. We do so by conducting two experimental studies. In our production task (Study I), we provide Dutch learners of Spanish with training on two phonemes that are typically difficult for them: /u/ and /θ/. We have chosen to approach L2 phoneme acquisition within the context that will likely be typical for adult L2 learners: They usually learn the L2 in a classroom setting and, in contrast to infants, are generally able to read, which means they often receive a large part of their instruction from written textbooks and exercises and part of the challenge lies therefore in making the correct association between spelling and sound. This means that producing the right L2 phoneme is not only dependent on whether they are familiar with the sound itself but also on whether they are accustomed to relating that particular sound to the correct grapheme. Prior research (e.g., Escudero et al., 2014) has shown that stimuli with incongruent grapheme-orthography mapping hinder L2 performance in various areas. We employed this distinction in order to manipulate phoneme complexity in our study: While there are subtle phonetic differences between the production of /u/ in Dutch and Spanish, it is a segment that is present in both the Dutch and the Spanish phoneme inventory. The difficulty for Dutch learners of Spanish lies in the fact that, in Spanish, the phoneme that corresponds to the grapheme < u > is always /u/, whereas in Dutch several phonemes correspond to the grapheme < u > , for instance, /æ/ as in *dun* ("thin"), /y/ as in *pure* ("pure") and in combination with other vowels there is even more variation possible with realizations, for instance, as /æy/, /ø/, or /ɑu/, as in *muis* ("mouse"), *leuk* ("fun"), or *rauw* ("raw," Kooij and Van Oostendorp, 2003). Conversely, when it comes to the acquisition of /θ/, the challenge is 2-fold: not only is /θ/ not a part of the Dutch phoneme inventory[2] and thus a new segment for which a category needs to be created, its only corresponding grapheme in Spanish is <z>,[3] while in Dutch <z> is typically pronounced as /z/ or /s/. In sum, while /u/ requires a novel grapheme to phoneme correspondence, /θ/ requires both a novel grapheme to phoneme correspondence and the creation of a new category in the phoneme inventory. These differences between /u/ and /θ/ allow us to manipulate phoneme complexity in our production task.

Our Dutch learners of Spanish received instruction on /u/ and /θ/ in one of four conditions: audio-only (AO), audio-visual (AV), audio-visual with a pointing gesture (AV-P), or audio-visual with an iconic gesture (AV-I). The AO condition serves as a baseline, to which we will compare the other conditions, of which the latter two contain either a less or more complex gesture: A pointing gesture was chosen as a less complex gesture, as it has no intrinsic semantic meaning and only serves to draw the listeners attention to a specific feature in the context, in our case, the mouth of the native speaker of Spanish pronouncing an example item. An iconic gesture was chosen as a more complex gesture, as it does have intrinsic semantic meaning because it illustrates to the listener which articulator is involved in the production of the target sounds and in which way it should be used. Our analyses will focus on whether gesture complexity and phoneme complexity affect the production of the target phonemes by Dutch learners of Spanish. In a perception task (Study II), Spanish natives listened to words containing the target phonemes that were produced by the Dutch learners of Spanish before and after AV, AV-P, or AV-I training and judged them on foreign accentedness and comprehensibility.

Based on previous studies (Hardison, 2003; Hazan et al., 2005), we hypothesize that adding audio-visual information to L2 phoneme training will facilitate phoneme acquisition, as compared to providing only audio information. Given that some previous work (e.g., Hannah et al., 2017; Kelly et al., 2017)

[2] While /θ/ is not included in the Dutch phoneme inventory, it is not the case that our participants are completely unfamiliar with this phoneme given that the participants of the current study, who are for the most part university students, will have all had at least 6 years of English education. However, /θ/ is a notoriously difficult phoneme for L1 speakers of Dutch in English as well (Gussenhoven and Broeders, 1997; Collins and Mees, 2003; Van den Doel, 2006; Hanulíková and Weber, 2012) and there are subtle differences in the phonetic realization of /θ/ in English (dental) and Spanish (interdental) that are still to be acquired.

[3] We are aware that the grapheme <c> is also pronounced as /θ/ in certain contexts in Castilian Spanish, but have chosen not to include this grapheme in our design in order to not overcomplicate the task for the novice Dutch learners of Spanish who participated in our production task.

has shown that gestures can be helpful in the acquisition of certain phonemes, we expect that including gestures in language training will be more beneficial than not including them, but possibly only in a context that is less cognitively demanding, that is, when producing /u/, but not /θ/ (Kelly and Lee, 2012). This would be in line with an embodied approach to cognition, which implies that not only performing but also seeing gestures benefits memory performance, which is essential in our phoneme production task (Madan and Singhal, 2012). Finally, given the lack of previous work that directly compares the role of different types of gestures in language acquisition, we cannot predict different effects between different types of gestures, but we speculate that there might be a difference between the potential facilitative effect of deictic and iconic gestures, based on the cognitive resources needed to process them. If this indeed affects their effectiveness in L2 pronunciation training, one would expect that pointing gestures might be more helpful than iconic gestures, which would be more cognitively demanding and thus entail less processing resources available for the perception and acquisition of the phoneme itself.

STUDY I

Method

Participants

In study I, 50 native speakers of Dutch, who did not speak any Spanish, took part. They were 28 women and 22 men, with a mean age of 25 years old (range 18–61 years old). Participants had no auditory or visual impairments that could affect their participation. Participants were recruited via the Radboud University research participation system and received either credits or a small financial reward for taking part.

Design

Study I consisted of a pretest – training – posttest paradigm. We used a between-subjects design in which participants took part in one of four experimental training conditions: AO (n = 12), AV (n = 13), AV-P (n = 13), or AV-I (n = 12). The dependent variable was the pronunciation of the target phonemes, coded as either on-target or not.

Materials

Sentences

In the pretest and the posttest, participants read out loud 16 Spanish four-word sentences (in one of two randomized orders) that were easy to parse, half of which were experimental items. In each experimental item, the first syllable of the two-syllable noun in the sentence contained either /u/ or /θ/ (e.g., *La nube es blanca, la zeta es verde)*. Each of the two target phonemes occurred in four target words, for /u/: *muro, nube, ruta, suma*; for /θ/: *zeta, zorro, zueco, zumo*. The eight remaining filler items also contained the target phonemes, but at different positions within the words or the sentence. The filler items were not analyzed. The target phonemes were embedded in the four-word sentences and presented to participants one at a time on PowerPoint slides. Each written sentence was accompanied by a picture illustrating the meaning of the sentence (see **Figure 1**). This was done to make the task more interesting and to help participants understand the semantic meaning of the sentence.

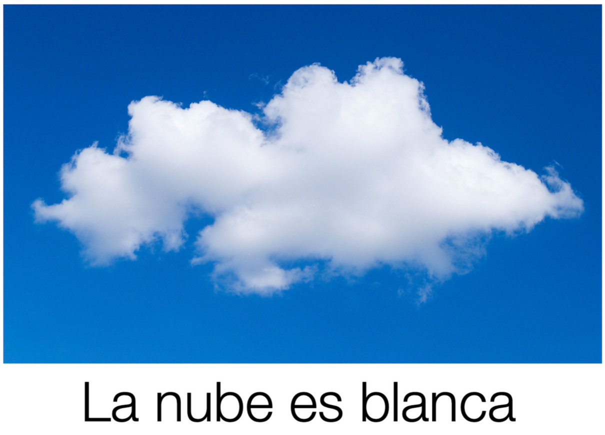
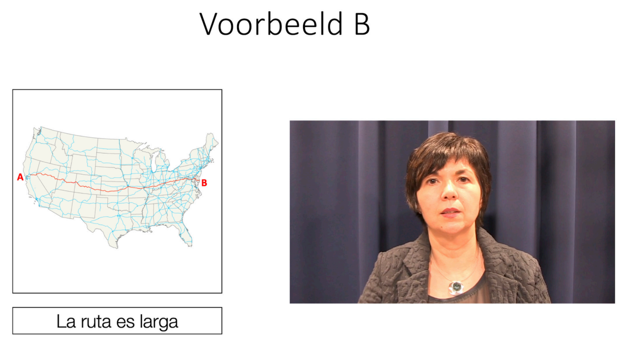


FIGURE 1 | Example of an experimental item containing the target phoneme /u/.

Training

After the pretest and before the posttest, participants received training on how to pronounce the target phonemes /θ/ and /u/ (in counterbalanced order) in Spanish. This training consisted of a set of three PowerPoint slides for each phoneme. On the first slide, written information was given on how to pronounce the target phoneme. Specifically, participants were told that the Spanish pronunciation of both graphemes differs from the Dutch pronunciation of these graphemes. Moreover, participants were explicitly instructed which articulatory gestures are necessary for nativelike pronunciation (i.e., "when pronouncing the letter "u" in Spanish, you need to round your lips" and "when pronouncing the letter "z" in Spanish, you need to place your tongue between your teeth and push out the air"). Apart from the written text, participants were also given an example of a native speaker of Spanish pronouncing the target phoneme in isolation. On the two following slides, participants were given two examples of the pronunciation of the target phoneme embedded within an example sentence. These examples (all produced by the same native speaker of Spanish) were accompanied by the written sentence and a picture illustrating the meaning of the sentence, in the same way as during the pretest and posttest (see **Figure 2**). The training was self-paced and participants took roughly 3 to 4 min to complete it. They were free to listen to/view the example fragments as many times as they wanted.

To manipulate training condition, the visual information given in the examples during the training varied, while the same audio was dubbed over all conditions. In the AO condition, participants heard the audio examples but did not see any video recordings of the speaker. In the AV condition, participants saw a video clip of the speaker producing the examples, but the speaker did not move her hands. In the AV-P condition, participants saw videos in which the speaker produced a pointing gesture toward her mouth while she produced the target phoneme.



Voorbeeld B

La ruta es larga

FIGURE 2 | Example of slide illustrating phoneme pronunciation within a sentence; screenshot of video on the right, sentence pronounced by the native speaker on the left, with accompanying picture.

In the AV-I condition, participants saw the speaker produce an iconic gesture while she produced the target phoneme (see **Figure 3** for examples). This iconic gesture represented the articulatory gesture needed for on-target phoneme production, as was explained verbally on the first training slide. For /u/, the iconic gesture was a one-handed gesture representing the rounding of the lips, and for /θ/, the iconic gesture was a one-handed gesture indicating that the speaker should push their tongue between their teeth. Both iconic gestures were made with one hand, roughly equally complex with respect to finger configuration, and not necessarily representing all articulators in the gesture but only the most relevant one for the learner. In the case of /θ/, Dutch learners of Spanish are familiar with non-sibilant fricatives (e.g., /f/ and /v/) but not interdental ones, so they need to know that they should push their tongue out of their mouth, which is only possible by placing it in between the teeth and lips. Concerning /u/, Dutch learners of Spanish need to know that correct pronunciation requires a stronger rounding of the lips than needed for any of the Dutch vowels. We performed a posttest for our stimuli among 42 native speakers of Dutch in which we compared the iconic and pointing gestures used for both phonemes with respect to how useful they found the gesture in the context of the L2 training for that specific phoneme, how intuitive they found the gesture in that context and whether they thought they understood why the gesture was chosen in that context. No significant differences were found between gesture type conditions or phoneme conditions for any of our measures, nor did the test reveal any significant interactions. This suggests that any differences between the iconic gestures concerning the way they visualize the relevant articulator did not affect our results.

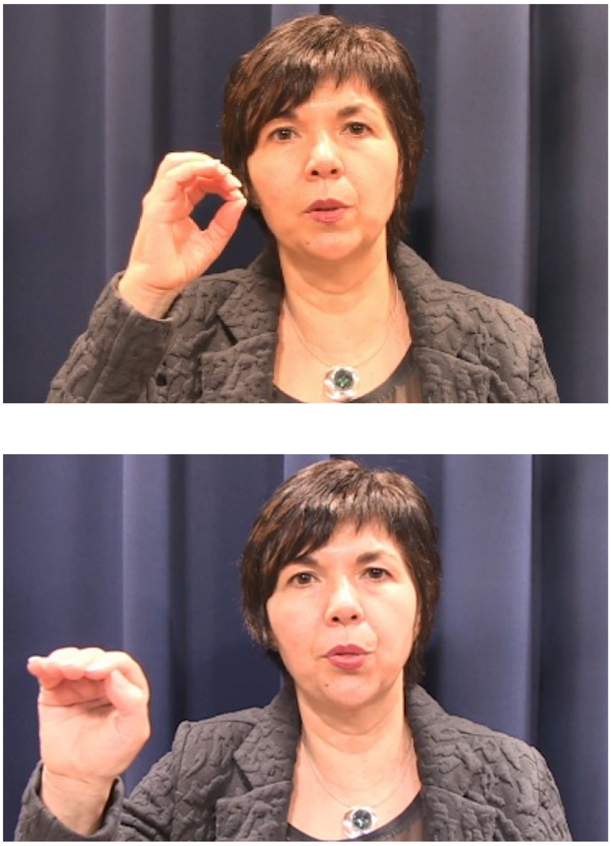


FIGURE 3 | Stills from training video in AV-I condition showing the articulatory gesture needed for /u/ (top still) and /θ/ (bottom still).

Procedure

To minimize distractions for the participants, the experiment took place in a soundproof booth. The language used throughout the experiment, except for the Spanish sentences during pretest, training, and posttest was Dutch. After participants had received instructions and signed a consent form, they were recorded while they read the 16 Spanish sentences out loud into a microphone (pretest). The pretest was first followed by a language background questionnaire, and then by one of the four types of pronunciation training. After the pronunciation training, participants were again recorded while they read out loud the same 16 Spanish sentences in a reordered version (posttest). Both the pretest and posttest were self-paced and participants were invited to repeat the sentences until they were satisfied with their pronunciation. The last production of each sentence was used for analysis. After completing all tasks, participants were debriefed.

Results

The audio recorded during the pretest and the posttest was annotated using Praat (Boersma and Weenink, 2018) concerning the production of the target phonemes. Two phonetically trained coders annotated the 1600 target phonemes (50 speakers × 16 sentences × 2 testing moments), and distinguished between a nativelike production (i.e., as a native speaker of Iberian Spanish would do) and several non-nativelike productions that are typical for native speakers of Dutch (for /θ/, these were /s/, /z/, or "other"; for /u/, these were /y/, /ə/, /ʏ/, or "other"). In the current analyses, nativelike productions were distinguished from the various non-nativelike productions, collapsing over the various non-target options. There was an overlap of 50% in coding and a good inter-rater reliability ($\kappa = 0.900$, $p < 0.001$).

TABLE 1 | Frequency of Training outcomes for /u/ and /θ/, separated by training condition.

Training Condition	Learning		Always Able		Never Able		Unlearning		Total
	/u/	/θ/	/u/	/θ/	/u/	/θ/	/u/	/θ/	
AO	**9**	**14**	36	0	0	32	0	0	91
AV	**16**	**19**	35	1	1	32	0	0	104
AV-P	**15**	**25**	32	0	2	26	2	0	102
AV-I	**21**	**10**	23	0	2	37	1	1	95
Total	**61**	**68**	126	1	5	127	3	1	392[4]

The target outcome is printed in bold.

Productions of target phonemes from the same sentences were compared between the pretest and the posttest, resulting in four different outcome options: (1) the participant was able to produce the target phoneme in the pretest, but not anymore at the posttest; (2) the participant was not able to produce the target phoneme at either the pretest or the posttest; (3) the participant was able to produce the target phoneme both at the pretest and at the posttest; (4) the participant was unable to produce the target phoneme at the pretest, but able to do so at the posttest. **Figure 4** and **Table 1** summarize the results per learning outcome separated by gesture condition and phoneme. In **Table 1**, the results are presented in terms of raw frequencies, while percentages are presented in **Figure 4**. First, we will inspect the data descriptively, followed by inferential statistics in the form of a mixed effects logistic regression analysis in which we distinguished between cases of "learning" (i.e., option 4), and "no learning" (i.e., collapsing options 1–3).

When inspecting the raw data per training condition in the cases that learning occurred, the Dutch learners of Spanish, in general, appear to benefit from receiving both auditory and visual information. For both phonemes, the cases of learning increase as more visual information is added, except for in the AV-I condition: While the L2 learners who aimed to produce a /u/ benefitted most from seeing an iconic gesture during training, the participants who aimed to produce a /θ/ appeared to benefit most from seeing a pointing gesture.

We used R (version 3.6.1, RCoreTeam, 2019) and the *lme4* package (Bates et al., 2015) to conduct a linear mixed effects logistic regression analysis to model binary outcome variables. The theoretically relevant predictors Gesture Condition (AO, AV, AV-P, or AV-I) and Phoneme (/u/ or /θ/) were included as fixed factors, and Training Outcome (Learning or No Learning) served as the response variable. Random intercepts were added for Speaker and Item. Adding random slopes resulted in models that either failed to converge or had inferior fit. Significance was assessed via likelihood ratio tests comparing the full model to a model lacking only the relevant effect. The complete model provided the best fit as determined by the Akaike Information Criterion, see **Table 2** for a complete overview of all effects and coefficients.

The analysis revealed that the condition that the participant was assigned to significantly predicted whether learning occurred or not but only when comparing the AV-I condition to the AO condition (the baseline condition, $\beta = 1.79$, $p < 0.05$). As gesture condition changes from AO to AV-I, the change in the odds of learning (rather than not learning) is 6.01. In other words, in general, a participant is more likely to learn than not in the AV-I condition than in the AO condition. In addition, there was an interaction between Phoneme and Gesture Condition ($\beta = -2.30$, $p < 0.01$), suggesting that the success of being in the AV-I condition depended on whether the participant aimed to produce a /u/ or a /θ/. The odds ratio tells us that as the gesture condition changes from AO to AV-I in combination with the phoneme being produced being a /θ/ instead of a /u/, the change in the odds of learning compared to not learning was 0.10. In order words, as the phoneme that is produced is /θ/ instead of /u/, participants are less likely to learn in the AV-I condition.

Interim Discussion

In summary, Study I showed that, in general, adding audio-visual information to phoneme pronunciation training aided target-like production. However, the complexity of the gesture produced by the trainer in combination with the complexity of the target phoneme affected L2 learners' success. Only when producing the less complex phoneme /u/, did participants benefit from seeing a more complex, iconic, gesture, making the AV-I condition the one in which L2 learners were most likely to learn. Conversely, when aiming to produce the more challenging phoneme /θ/, seeing a more complex gesture was actually detrimental to L2 learners, resulting in less learning taking place than in all other conditions. Additionally, the analysis corroborates our theoretical predictions concerning the complexity level of both phonemes. L2 learners often tended to already produce /u/ in a target-like way during the pretest, whereas they generally continued to be unable to correctly produce /θ/ during the posttest. This confirms that /u/ inherently is a less complex phoneme for Dutch learners of Spanish than /θ/.

STUDY II

Method

Participants

For this study, the data of 103 Spanish natives was analyzed. They were from the center of Spain; either from the autonomous region of Madrid, Castilla-La Mancha, or Castilla-León. On average, they were 30.9 years old (*SD* = 6.6 years) and 52 of them were women. None of the participants had auditory impairments that could have affected participation in the experiment. Participants were recruited via Qualtrics (2020) and received a small monetary reward for their participation.

[4] In theory, 400 comparisons can be made between the performance at pretest and posttest (4 items × 2 segments × 50 participants × 2 moments = 800 productions, which equals 400 comparisons). In practice, 8 data points were lost due to inferior sound quality or coding difficulties. Per training condition, the maximally possible total is 96 for AO and AV-I and 104 for AV and AV-P. For relative frequencies in the form of percentages, see **Figure 4**.

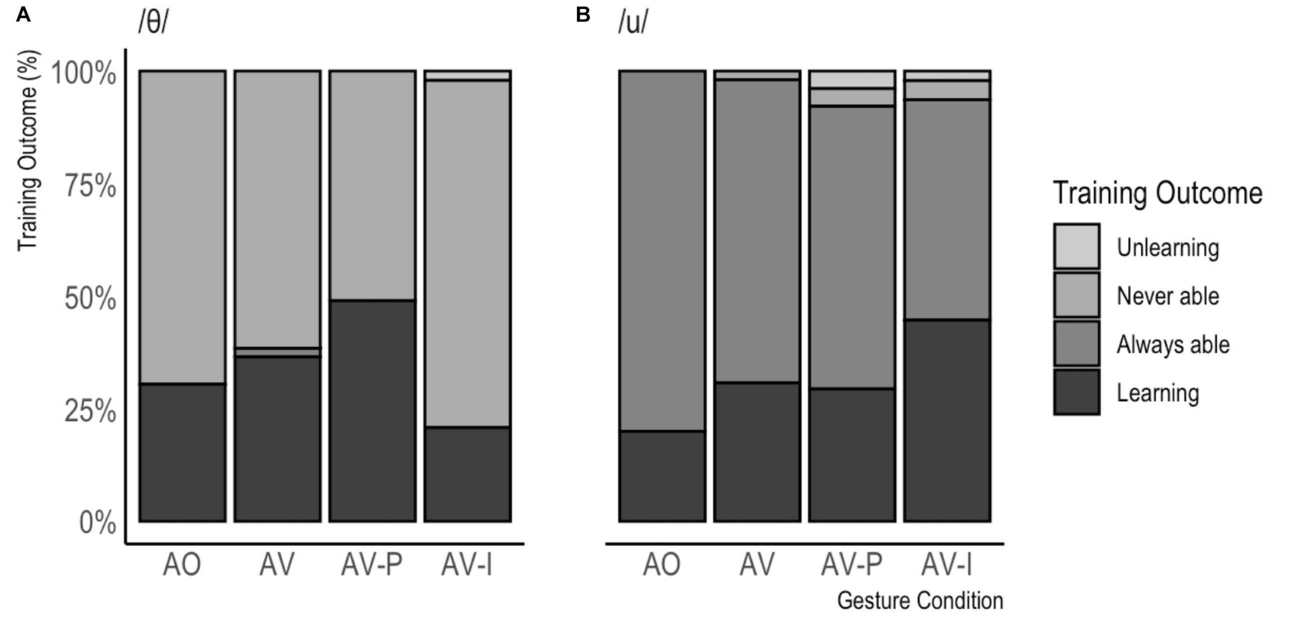


FIGURE 4 | Training Outcome in percentages, separated by Gesture Condition for /θ/ **(A)** and /u/ **(B)**.

TABLE 2 | Estimated effects and coefficients for Training Outcome.

Learning vs. Not Learning	**β estimate**	**Std. error**	***z* value**	***p* value**	**95% CI for Odds Ratio**		
					Lower	**Odds Ratio**	**Upper**
Intercept	**−2.06**	**0.68**	**−3.06**	**0.002**	**0.03**	**0.13**	**0.48**
Gesture Condition$_{AV}$	0.86	0.77	1.12	0.263	0.52	2.37	10.74
Gesture Condition$_{AV-P}$	0.90	0.76	1.19	0.236	0.55	2.47	11.01
Gesture Condition$_{AV-I}$	**1.79**	**0.77**	**2.32**	**0.020**	**1.32**	**6.01**	**27.33**
Phoneme$_{/\theta/}$	0.88	0.73	1.20	0.232	0.57	2.41	10.14
Phoneme$_{/\theta/}$ * Gesture Condition$_{AV}$	−0.44	0.76	−0.59	0.558	0.15	0.64	2.82
Phoneme$_{/\theta/}$ * Gesture Condition$_{AV-P}$	0.23	0.73	0.32	0.753	0.30	1.26	5.30
Phoneme$_{/\theta/}$ * Gesture Condition$_{AV-I}$	**−2.30**	**0.78**	**−2.95**	**0.003**	**0.02**	**0.10**	**0.46**

Random effects	**Variance**	**Standard deviation**
Speaker	1.593	1.262
Item	0.423	0.650

The intercept represents the following combination of variable levels: Gesture Condition = AO, Phoneme = /u/. Asterisks (*) represent interactions, subscript signals the level of a categorical variable. Significant p-values are printed in bold. The model used in this analysis can be described as Training Outcome ∼ Gesture Condition × Phoneme + (1| participant) + (1| item).

Design

The experiment had a within-subjects design in which participants listened to target words from the pretest and the posttest produced by a subset of the participants of Study I. The productions were taken from three out of the four experimental conditions of Study I: AV, AV-P, and AV-I. The AO condition was left out to reduce the length of the perception tasks for the participants and because it represented a less natural learning context; most L2 learners are taught in a classroom setting where they can see the teacher. All participants judged these words for both perception measures.

Materials

Participants listened to randomly ordered target words produced by participants of study I. Because it was not feasible to have participants in study II to listen to all target words from experiment I, a selection was made. We used 8 items (2 with /u/ and 2 with /θ/, from both the pretest and the posttest) from 21 randomly selected speakers of study I. To make the experiment as interesting as possible for participants, the selected items were not the same ones for every speaker (e.g., the productions of *muro, nube, zeta,* and *zorro* as produced during the pretest and posttest were taken from one speaker, and productions of *ruta, suma, zueco,* and *zumo* as produced during the pretest and posttest were taken from another speaker). In total, participants listened to 168 items per measure (7 speakers × 2 items per phoneme × 2 phonemes × 3 gesture conditions × 2 testing moments).

Instruments

In separate blocks, participants judged each of the 168 items on accentedness and comprehensibility. Based on Derwing and Munro (1997), accentedness was measured with the statement "This person speaks . . .", followed by a 7-point semantic differential anchored by "with a strong foreign accent – without a strong foreign accent" and comprehensibility was measured with the statement "This person is. . .", followed by a 7-point semantic differential anchored by "Very hard to understand – very easy to understand".

Procedure

The entire experiment took place online, in Spanish. Participants were given information about the experiment, and a consent form to sign, after which they filled out a short questionnaire on their language background. Subsequently, they performed the Spanish LexTALE task (Izura et al., 2014), which is a measure of Spanish vocabulary size. This enabled us to check that they were taking the task seriously, because, as native speakers, they should all be able to generate a high LexTALE score. Hence, any participants that performed below the L1 threshold of 47 points in the test were excluded from the final analysis. In the main part of the experiment, participants were randomly assigned to a block to rate one measure (either accentedness or comprehensibility), followed by a block rating the other measure. The entire experiment took about 30 min to complete.

Results

Using R and the *psych* package (Revelle, 2019), the intra-class correlation coefficient was computed to assess the agreement between participants in rating the accentedness and comprehensibility of the words produced by our Dutch learners of Spanish in Study I. For both accentedness and comprehensibility, there was an excellent absolute agreement between the participants, using the two-way random effect models and "single rater" unit, both $\kappa = 0.94$, $p < 0.05$, which implies that they strongly agreed amongst themselves in regards to the accentedness and comprehensibility of the speech fragments that they listened to. In what follows, we will first report the data descriptively, followed by a report of the inferential statistics in the form of ordinal regression analyses for both measures.

The accentedness ratings per gesture condition (separated by phoneme) are visualized in **Figure 5** and **Table 3** contains the descriptive statistics per testing moment and gesture condition (split by phoneme), for accentedness. As can be seen from these results, the effects found in production appear to hold for perception as well: For items containing /u/, the difference between pre- and posttest is largest in the AV-I condition, whereas, for items containing /θ/, this difference is virtually non-existent in the AV-I condition, but largest in the AV-P condition.

TABLE 3 | Accentedness ratings per predictor for items containing /u/ and /θ/.

Testing Moment	Gesture Condition	/u/		/θ/	
		Mean	***Standard deviation***	***Mean***	***Standard deviation***
Pretest	AV	3.39	1.81	3.58	1.74
	AV-P	3.72	1.81	3.63	1.69
	AV-I	3.48	1.76	3.45	1.77
Posttest	AV	3.58	1.88	3.53	1.75
	AV-P	3.84	1.78	3.95	1.83
	AV-I	3.82	1.89	3.47	1.79

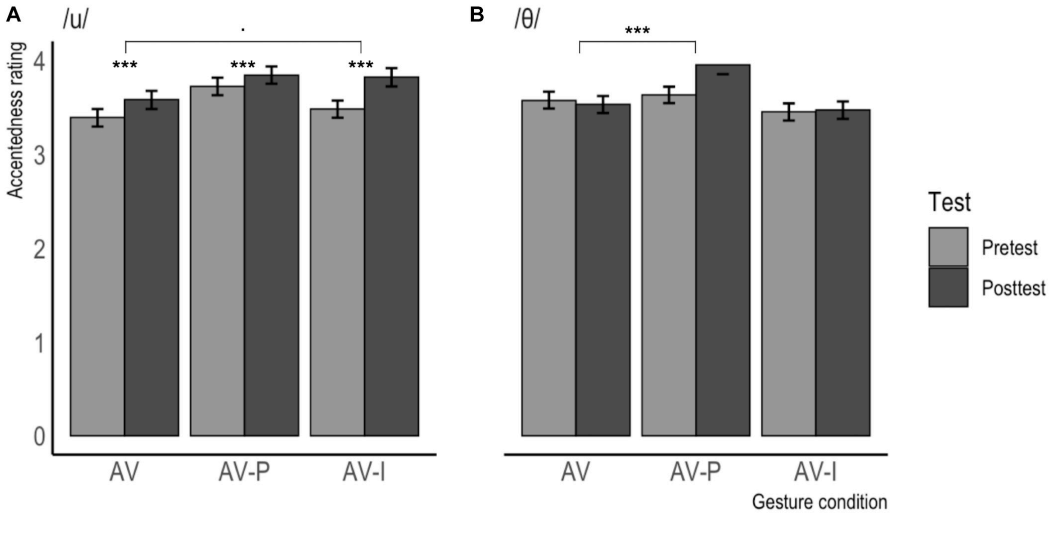


FIGURE 5 | Mean accentedness ratings for /u/ **(A)** and /θ/ **(B)** produced at Pretest and Posttest, separated by gesture condition. Error bars represent confidence intervals. Higher scores indicate a less strong foreign accent.

The comprehensibility ratings per gesture condition (separated by phoneme) are visualized in **Figure 6** and **Table 4** contains the descriptive statistics per testing moment and gesture condition, split by phoneme. In general, it can be noted that the comprehensibility ratings are roughly one scale point higher than the accentedness ratings. For items containing /u/, the difference between pre- and posttest is again largest in the AV-I condition, whereas for items containing /θ/, speakers who were in the AV-I condition were judged more difficult to comprehend after training than before, while they were deemed slightly easier to understand after training if they had been in the AV-P condition.

We will now evaluate the statistical evidence for the findings described above, which were based on visual inspection. We fitted ordinal regression models with random effects on the data for accentedness and comprehensibility separately, using R and the clmm function from the *ordinal* package (version 12-10, Christensen, 2019). We included the theoretically relevant predictors in the model: Testing Moment (pretest or posttest), Gesture Condition (AV, AV-P, or AV-I), Phoneme (/u/ or /θ/), and random intercepts by Participant, Speaker, and Item. Adding random slopes resulted in models that either failed to converge or had inferior fit. Significance was assessed via likelihood ratio tests comparing the full model to a model lacking only the relevant effect. The complete model provided the best fit as determined by the Akaike Information Criterion, see **Tables 5**, **6**.

TABLE 4 | Comprehensibility ratings per predictor for items containing /u/ and /θ/.

Testing Moment	Gesture Condition	/u/		/θ/	
		Mean	*Standard deviation*	*Mean*	*Standard deviation*
Pretest	AV	4.44	1.81	4.54	1.76
	AV-P	4.68	1.77	4.66	1.68
	AV-I	4.48	1.77	4.40	1.76
Posttest	AV	4.56	1.81	4.47	1.77
	AV-P	4.78	1.74	4.75	1.80
	AV-I	4.69	1.80	4.28	1.86

Accentedness

The ordinal regression analysis for accentedness revealed no main effects of Testing Moment, Gesture Condition or Phoneme, but several significant interactions were found, see **Table 5**. The analysis revealed a significant interaction effect between Testing Moment and Gesture Condition, with a bigger difference between the ratings at Pretest and Posttest in the AV-P condition (Pretest: $M = 3.68$, $SD = 1.75$; Posttest: $M = 3.89$, $SD = 1.81$; $M\Delta = 0.21$) than in the AV condition (Pretest: $M = 3.48$, $SD = 1.78$; Posttest: $M = 3.55$, $SD = 1.81$; $M\Delta = 0.07$). In addition, a significant interaction was found between Testing Moment and Phoneme, with a bigger difference between the ratings at Pretest and Posttest for the items containing /u/ (Pretest: $M = 3.53$, $SD = 1.80$; Posttest: $M = 3.75$, $SD = 1.85$; $M\Delta = 0.22$) than for those containing /θ/ (Pretest: $M = 3.55$, $SD = 1.74$; Posttest: $M = 3.65$, $SD = 1.80$; $M\Delta = 0.10$). The analysis also revealed a significant interaction effect between Gesture Condition and Phoneme, with a bigger difference between the AV and AV-P conditions for the items containing /u/ (AV: $M = 3.48$, $SD = 1.84$; AV-P: $M = 3.78$, $SD = 1.80$; $M\Delta = 0.30$) than for those containing /θ/ (AV: $M = 3.55$, $SD = 1.75$; AV-P: $M = 3.79$, $SD = 1.77$; $M\Delta = 0.24$).

Finally, the model revealed a three-way interaction between Testing Moment, Phoneme, and Gesture condition. In order to interpret this interaction, we performed two separate mixed

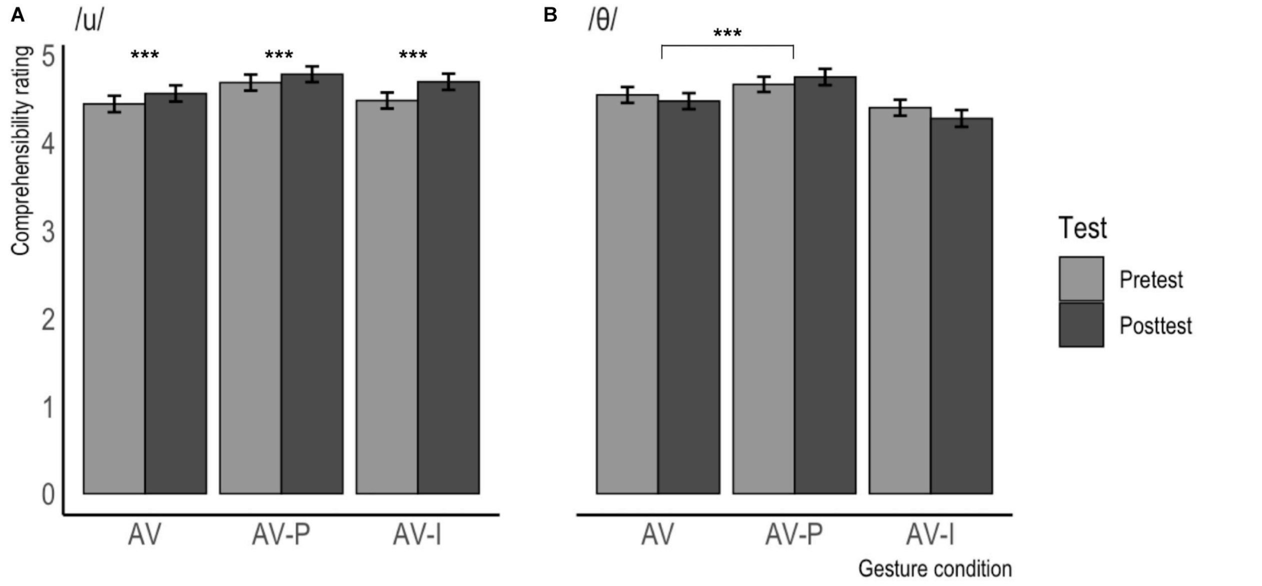


FIGURE 6 | Mean comprehensibility ratings for /u/ **(A)** and /θ/ **(B)** produced at pretest and posttest, separated by gesture condition. Error bars represent confidence intervals. Higher scores indicate higher comprehensibility.

TABLE 5 | Estimated effects and coefficients for accentedness ratings.

Predictor	β estimate	Std. error	*z* value	*p* value
Testing Moment$_{\text{POSTTEST}}$	−0.055	0.066	−0.830	0.406
Gesture Condition$_{\text{AV−P}}$	0.066	0.223	0.295	0.768
Gesture Condition$_{\text{AV−I}}$	−0.058	0.225	−0.256	0.798
Phoneme$_{\text{/u/}}$	−0.139	0.356	−0.390	0.696
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−P}}$	**0.428**	**0.093**	**4.478**	**0.000**
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−I}}$	0.086	0.093	0.917	0.359
Testing Moment$_{\text{POSTTEST}}$ * Phoneme$_{\text{/u/}}$	**0.284**	**0.094**	**3.018**	**0.003**
Gesture Condition$_{\text{AV−P}}$ * Phoneme$_{\text{/u/}}$	**0.337**	**0.093**	**3.608**	**0.000**
Gesture Condition$_{\text{AV−I}}$ * **Phoneme**$_{\text{/u/}}$	0.050	0.094	0.527	0.598
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−P}}$ * Phoneme$_{\text{/u/}}$	**−0.492**	**0.133**	**−3.704**	**0.000**
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−I}}$ * **Phoneme**$_{\text{/u/}}$	0.074	0.133	0.559	0.576

Random effects	Variance	Standard deviation
Participant	1.225	1.107
Speaker	0.159	0.399
Item	0.245	0.495

The intercept represents the following combination of variable levels: Testing Moment = Pretest, Gesture Condition = AV, Phoneme = /θ/. Asterisks (*) represent interactions, subscript signals the level of a categorical variable. Significant p-values are printed in bold. The model used in this analysis can be described as: Rating ∼ Testing Moment × Gesture Condition × Phoneme + (1| Participant) + (1 | Speaker) + (1 | Item).

TABLE 6 | Estimated effects and coefficients for comprehensibility ratings.

Predictor	β estimate	Std. error	*z* value	*p* value
Testing Moment$_{\text{POSTTEST}}$	−0.098	0.066	−1.480	0.139
Gesture Condition$_{\text{AV−P}}$	0.093	0.215	0.434	0.664
Gesture Condition$_{\text{AV−I}}$	−0.070	0.216	−0.324	0.746
Phoneme$_{\text{/u/}}$	−0.012	0.363	−0.034	0.973
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−P}}$	**0.264**	**0.094**	**2.802**	**0.005**
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−I}}$	−0.031	0.094	−0.331	0.740
Testing Moment$_{\text{POSTTEST}}$ * Phoneme$_{\text{/u/}}$	**0.262**	**0.094**	**2.783**	**0.005**
Gesture Condition$_{\text{AV−P}}$ * Phoneme$_{\text{/u/}}$	**0.216**	**0.094**	**2.305**	**0.021**
Gesture Condition$_{\text{AV−I}}$ * **Phoneme**$_{\text{/u/}}$	−0.004	0.094	−0.046	0.963
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−P}}$ * Phoneme$_{\text{/u/}}$	**−0.313**	**0.134**	**−2.335**	**0.020**
Testing Moment$_{\text{POSTTEST}}$ * Gesture Condition$_{\text{AV−I}}$ * Phoneme$_{\text{/u/}}$	0.128	0.133	0.962	0.336

Random effects	Variance	Standard deviation
Participant	1.756	1.325
Speaker	0.147	0.383
Item	0.255	0.505

The intercept represents the following combination of variable levels: Testing Moment = Pretest, Gesture Condition = AV, Phoneme = /θ/. Asterisks (*) represent interactions, subscript signals the level of a categorical variable. Significant p-values are printed in bold. The model used in this analysis can be described as: Rating ∼ Testing Moment × Gesture Condition × Phoneme + (1| Participant) + (1| Speaker) + (1| Item).

ordinal regression analyses, one on items containing /u/ and one on items containing /θ/. These analyses show that the above-mentioned interaction between Testing Moment$_{\text{POSTTEST}}$ and Gesture Condition$_{\text{AV−P}}$ was significant for the items containing /θ/ ($\beta = 0.434$, $SE = 0.094$, $z = 4.628$, $p < 0.001$), but not significant for the items containing /u/ ($\beta = -0.062$, $SE = 0.094$, $z = -0.660$, $p = 0.509$). For items containing /θ/, there was a bigger difference between the ratings at Pretest and Posttest in the AV-P condition (Pretest: $M = 3.63$, $SD = 1.69$; Posttest: $M = 3.95$, $SD = 1.83$; $M\Delta = 0.32$) than in the AV condition (Pretest: $M = 3.58$, $SD = 1.74$; Posttest: $M = 3.53$, $SD = 1.75$; $M\Delta = -0.05$). For items containing /u/, there was no difference between the ratings at Pretest and Posttest in the AV-P condition (Pretest: $M = 3.72$, $SD = 1.81$; Posttest: $M = 3.84$, $SD = 1.78$; $M\Delta = 0.12$) and the AV condition (Pretest: $M = 3.39$, $SD = 1.81$; Posttest: $M = 3.58$, $SD = 1.88$; $M\Delta = 0.19$). In addition, the analysis on the items containing /u/ revealed a trend for the interaction between Testing Moment$_{\text{POSTTEST}}$ and Gesture Condition$_{\text{AV−I}}$ ($\beta = 0.176$, $SE = 0.095$, $z = 1.856$, $p = 0.063$) in which there was a bigger difference between the ratings at Pretest and Posttest in the AV-I condition (Pretest: $M = 3.48$, $SD = 1.76$; Posttest: $M = 3.82$, $SD = 1.89$; $M\Delta = 0.34$) than in the AV condition

(Pretest: M = 3.39, SD = 1.81; Posttest: M = 3.58, SD = 1.88; $M\Delta$ = 0.19). Finally, the analysis on the items containing /u/ also revealed a significant main effect of Testing Moment, with higher ratings at Posttest (M = 3.75, SD = 1.85) than at Pretest (M = 3.53, SD = 1.80), irrespective of Gesture Condition.

Comprehensibility

For comprehensibility, the ordinal regression analysis also revealed no significant main effects, but several significant interactions between the fixed factors were found. The analysis revealed a significant interaction effect between Testing Moment and Gesture Condition, with a bigger difference between the ratings at Pretest and Posttest in the AV-P condition (Pretest: M = 4.67, SD = 1.72; Posttest: M = 4.76, SD = 1.77; $M\Delta$ = 0.09) than in the AV condition (Pretest: M = 4.49, SD = 1.79; Posttest: M = 4.52, SD = 1.79; $M\Delta$ = 0.03). In addition, a significant interaction was found between Testing Moment and Phoneme, with a bigger difference between the ratings at Pretest and Posttest for the items containing /u/ (Pretest: M = 4.53, SD = 1.79; Posttest: M = 4.68, SD = 1.78; $M\Delta$ = 0.15) than for those containing /θ/ (Pretest: M = 4.53, SD = 1.74; Posttest: M = 4.50, SD = 1.82; $M\Delta$ = −0.03). The analysis also revealed a significant interaction effect between Gesture Condition and Phoneme, with a bigger difference between the AV and AV-P conditions for the items containing /u/ (AV: M = 4.50, SD = 1.81; AV-P: M = 4.73, SD = 1.76; $M\Delta$ = 0.23) than for those containing /θ/ (AV: M = 4.51, SD = 1.76; AV-P: M = 4.71, SD = 1.74; $M\Delta$ = 0.20).

Finally, the model revealed a three-way interaction between Testing Moment, Phoneme, and Gesture condition. In order to interpret this interaction, we performed two separate mixed ordinal regression analyses, one on items containing /u/ and one on items containing /θ/. These analyses show that the above-mentioned interaction between Testing Moment$_{\mathrm{POSTTEST}}$ and Gesture Condition$_{\mathrm{AV-P}}$ was significant for the items containing /θ/ (β = 0.268, SE = 0.094, z = 2.843, $p < 0.01$) but not significant for the items containing /u/ (β = −0.037, SE = 0.095, z = −0.384, p = 0.701). For items containing /θ/, there was a bigger difference between the ratings at Pretest and Posttest in the AV-P condition (Pretest: M = 4.66, SD = 1.68; Posttest: M = 4.75, SD = 1.80; $M\Delta$ = 0.09) than in the AV condition (Pretest: M = 4.54, SD = 1.76; Posttest: M = 4.47, SD = 1.77; $M\Delta$ = −0.07). For items containing /u/, there was no difference between the ratings at Pretest and Posttest in the AV-P condition (Pretest: M = 4.68, SD = 1.77; Posttest: M = 4.78, SD = 1.74; $M\Delta$ = 0.10) and in the AV condition (Pretest: M = 4.44, SD = 1.81; Posttest: M = 4.56, SD = 1.81; $M\Delta$ = 0.12). In addition, the analysis on the items containing /u/ revealed a significant main effect of Testing Moment, with higher ratings at Posttest (M = 4.68, SD = 1.78) than at Pretest (M = 4.53, SD = 1.79), irrespective of Gesture Condition.

Interim Discussion

In summary, Study II showed that the findings of Study I, in which the more complex iconic gesture facilitated the production of the less complex phoneme /u/ but not the production of the more complex phoneme /θ/, and the less complex pointing gesture facilitated the production of the more complex phoneme /θ/ but less so for the production of the less complex phoneme /u/, were confirmed. When a pointing gesture was included in the training, this was particularly helpful for items containing /θ/, but not for items containing /u/, resulting in less foreign-accentedness and higher perceived comprehensibility for /θ/ items. For items containing /u/, seeing an iconic gesture during training lead to speech being judged as less foreign-accented but equally comprehensible. Also, again in line with the findings from Study I, Study II showed that /u/ was easier and /θ/ was harder to acquire; scores on foreign-accentedness and perceived comprehensibility differed more between the pretest and posttest for /u/ than for /θ/. Although these results show that the interaction between type of gesture and type of phoneme during training affects perceived accentedness and comprehensibility, we should realize that the effects were relatively small; the differences in scores between pretest and posttest were generally less than one point on a 7-point scale. Finally, we found that the ratings for accentedness were lower than the ratings for comprehensibility. As found in previous work, it appears that although native listeners are sensitive to hearing deviations from native pronunciation, this does not necessarily result in a lower comprehensibility score (Derwing and Munro, 1997; Munro and Derwing, 1999; Van Maastricht et al., 2016, 2020).

GENERAL DISCUSSION

The goal of this study was to investigate if gestures can facilitate L2 phoneme acquisition, and, more specifically, in what way the complexity of the gesture and the complexity of the phoneme play a role in this process. We focused on the acquisition of two Spanish phonemes which are typically hard for native speakers of Dutch: /u/ and /θ/. We expected /u/ to be easier to acquire because, although the grapheme it is typically associated with in Dutch differs from the grapheme typically used in Spanish, the phoneme /u/ does also occur in the Dutch phoneme inventory. We expected /θ/ to be harder to acquire because, in addition to the Spanish grapheme associated with this phoneme being pronounced differently in Dutch, the phoneme is not part of the Dutch phoneme inventory. We hypothesized that adding audio-visual information to the phoneme training that Dutch learners of Spanish received would facilitate phoneme acquisition, as compared to providing only audio information. In addition, we expected that including gestures in the phoneme training would be most beneficial for phonemes that are less cognitively demanding, in this case /u/, rather than /θ/. Phoneme training took place in one of four conditions: audio-only, audio-visual, audio-visual with a pointing gesture, or audio-visual with an iconic gesture. Given the lack of previous studies comparing the effect of different types of gestures on phoneme acquisition, we did not have clear predictions concerning which type of gesture would work best. Based on the idea that processing a pointing gesture is less cognitively demanding than processing an iconic gesture, we speculated that a pointing gesture might be more helpful than an iconic gesture during phoneme acquisition because processing a cognitively less demanding pointing gesture would leave more processing resources available for the perception and acquisition of the new phoneme, as compared to processing a cognitively more demanding iconic gesture. We conducted two studies to investigate these issues: Study I, in which native speakers of Dutch received training in one of the four conditions and produced the Spanish target phonemes in a pretest and posttest, and Study II, in which native speakers of Spanish listened to the words containing the target phonemes as produced by the Dutch learners of Spanish before and after training and scored these on accentedness and comprehensibility.

The results of both studies showed that, in general, adding audio-visual information to phoneme pronunciation training facilitates target-like production. However, it matters which gesture is added to the training of which phoneme, as the specific gesture-phoneme combination can result in more, or less, target-like production, accentedness, and perceived comprehensibility. Also, the results of both studies complement each other in the sense that the improvements in phoneme production in certain experimental conditions in Study I were generally reflected in less foreign-accentedness and higher comprehensibility ratings for items from these same conditions in Study II.

Returning to our hypotheses, we find that our data confirm our first prediction, namely that /u/ would be easier to acquire than /θ/ for Dutch learners of Spanish. Study I showed that /u/ was often already produced in a target-like manner during the pretest, whereas /θ/ was often never produced in a target-like manner, regardless of training condition. Study II also showed that between the pretest and posttest items containing /u/ were rated as less foreign-accented and more comprehensible, which was not the case for items containing /θ/. These findings suggest that if /u/ had not already been acquired before training, it can be acquired during training, but that in many cases, a single training session is not sufficient to benefit the acquisition of /θ/.

With respect to our second hypothesis, we find partial corroboration of earlier work (Hardison, 2003; Hazan et al., 2005) in our results. Study I revealed that adding audio-visual information to training that includes an iconic gesture affects target-like production, as compared to providing audio-only information. Whether this effect on target-like production is positive or negative depends on the phoneme in question: If the phoneme being acquired was /u/, seeing an iconic gesture during training led to more cases of learning. However, if the phoneme being acquired was /θ/, seeing an iconic gesture during training was detrimental, leading to fewer cases of learning than in all other conditions. In other words, seeing a complex gesture during training facilitated the target-like production of the easy phoneme, whereas seeing a complex gesture during training harmed the target-like production of the complex phoneme. The importance of the phoneme-gesture combination is also reflected in the results of Study II: The specific gesture being used during training could result in less foreign-accentedness and higher comprehensibility, but this depended on which phoneme was being produced. Seeing a less complex pointing gesture during training lead to productions of words with /θ/ that were perceived as less foreign-accented and more comprehensible. Seeing a more complex iconic gesture during training lead to productions of words with /u/ that were perceived as less foreign-accented. This means that our speculation, based on findings by Kelly et al. (2017), that the less cognitively demanding pointing gesture would facilitate acquisition most, was not supported by all the data: The facilitative effect of the pointing gesture depended on which phoneme was being acquired: pointing gestures worked best for the complex phoneme /θ/, but not for the easy phoneme /u/. The more complex iconic gesture helped in the acquisition of the easier phoneme /u/, but hindered acquisition of /θ/.

These results mean that the complexity of both the gesture and the phoneme matters when using gesture in L2 phoneme acquisition. It appears that a complex phoneme is best combined with a simple gesture, and the other way around. Most previous studies did not take the complexity of the target phoneme or gesture into account, and this may help to explain some of the contradictory prior findings. For instance, Kelly et al. (2014) investigated the effect of metaphoric gestures in the context of Japanese vowel length contrasts as perceived by American learners. While their study mainly revealed no differences between the learners that had seen or seen and produced gestures during training, they reported one case in which there was a significant difference between their experimental groups. While half of their participants were trained on metaphoric gestures representing the vowel length as a syllable, the other half was trained on gestures representing the vowel length as a mora. Reaction times of participants from the latter group were significantly longer than those of participants in the former group during an auditory identification test, which implies that American learners needed more time to process the gesture related to the mora category, which is non-native to them, than the gesture related to the syllable category, which is native to them. This is in line with the findings of the current study, because a gesture representing an unfamiliar phonemic element (mora) is arguably more complex for L2 learners than a gesture representing a familiar phonemic element (syllable).

Moreover, the current results are in line with the idea proposed by Kelly and Lee (2012) that gesture may only help when the task demands are not too high. In their work, they focused on L2 vocabulary learning and distinguished between phonetically easy and phonetically hard word pairs. Their results showed that (iconic) gestures helped for the easy word pairs, but actually hindered the vocabulary acquisition for the hard word pairs. Importantly, in their work, they did not distinguish between types of gestures but only used iconic gestures, and with the results they found they wonder whether it may be the case that gestures that convey less or no semantic meaning, as compared to iconic gestures, would also hinder acquisition, or whether the fact that iconic gestures carry semantic information is a reason for the fact that they do not always help, and may even hinder the learner. The pointing gesture used in our current study is an example of such a gesture that conveys less semantic information. After all, a pointing gesture mainly serves as a manual highlighter and, at least in the current study, does not provide any information about the speech it accompanies. If we contrast this with the iconic gestures used in our study we can see that those contained quite a bit of semantic information, more specifically, the gestures visualized what the relevant articulators were of the specific phoneme, and how these articulators should be used to produce the phoneme. This suggests that seeing the iconic gesture cost a fair amount of processing energy, as compared to seeing the pointing gesture. This processing energy may come at a cost to the resources that are left for focusing on listening to the sound of the phoneme and watching the actual articulators needed for phoneme production. If more cognitive energy is needed because the phoneme in question does not exist in the native language phoneme inventory, this may result in less, rather than more, acquisition taking place. Likewise, if the processing of the phoneme takes less cognitive effort, for example, because the phoneme is already familiar, there is more processing space left to take the gesture that is being produced into account. Again, this is in line with the suggestion given by Kelly and Lee (2012) that gesture may only facilitate the processing of sounds that are familiar in someone's native language.

Naturally, our findings can be expanded on in several ways. More studies are needed to further determine which elements of a gesture or phoneme contribute to its complexity. Specifically, it would be interesting to compare different types of gestures on only one dimension. The complex gestures used for the two phonemes in our study were both iconic in nature, but the one visualizing the articulator needed to produce /u/ highlighted the lips, whereas the one representing the articulator involved in the production of /θ/ highlighted the tongue. Comparing two phonemes that are more similar in their articulation but different with respect to their presence in the L1 inventory, such as /a/ and /ɑ/ for a native speaker of Spanish, would enable a comparison of two iconic gestures that represent the same articulators and that thus are more similar in form. Of course, this also generates a challenge: if the articulation of two phonemes is very similar, can the two gestures reflect the relevant information and remain sufficiently different to be useful? In the same vein, more different types of (less complex) gestures should be compared, for instance, beats versus pointing gestures. It might well be the case that gesture complexity is not so much a dichotomous concept, but rather one that spans a continuum.

Similarly, we have defined phoneme complexity in our study as the extent to which our participants were familiar with the used phonemes in their L1, but it seems reasonable to assume that other factors contribute to a phoneme's complexity. A comparison of two phonemes that are both not included in the participants' L1 inventory, but that differ in the necessary articulators, might generate more insight in whether phoneme complexity, like gesture complexity, is a continuum on which "presence in the L1 inventory" might be of more importance than "familiarity with the articulators." For instance, one could compare the uvular /χ/ and glottal /ɦ/, both of which are typical of Dutch but do not occur in the French phoneme inventory. While the French phoneme inventory does contain another uvular phoneme, /ʁ/, there are no other glottals in the system, which might make /ɦ/ a more difficult sound to acquire for French natives than /χ/. In addition, the effect of gesture and sound complexity might not only hold for segmental sounds but could also apply to suprasegmentals, as implied by the results of Kelly et al. (2017). Finally, the relative weight of certain segments in communication between L1 and L2 speakers may also be of consideration in this respect. As shown for English by Suzukida and Saito (2019), the Functional Load principle (as applied to L2 pronunciation teaching by Brown, 1988) can be used as a tool to determine which segments are crucial for successful understanding in L1-L2 communication.

A potential limitation to the current study is that participants in Study I received only one short training, and were tested almost immediately after this training. This means that we do not know to what extent the current results also apply to long term learning and whether repeated training yields different results. Results obtained by Zhen et al. (2019) and Li et al. (2020), which had similarly, short training sessions (of seven and two and a half minutes, respectively), imply that it is possible to obtain effects from only one short training, even long term. The fact that we found effects of gesture and phoneme complexity on the acquisition of L2 phonemes after only one short training corroborate their findings and are promising in the sense that we expect more or longer training to strengthen our results. Another potential limitation is that Study I took place in a laboratory setting, in which participants took part individually in a soundproof booth. Although this meant that we were able to control the experimental conditions and receive high-quality sound recordings, it also means that the external validity of this study is restricted, in the sense that the laboratory setting was not representative of a classroom setting in which pronunciation training may normally take place.

In conclusion, more research is needed in the context of the possibly beneficial role of gestures in foreign language acquisition and the role of complexity in this context. Prior, present, and future results in this context do not only further inform the theory regarding the nature of multimodal communication and (foreign) language learning, but are also directly relevant in practice. In (foreign) language acquisition, but also in many other fields, knowing which gesture works in which context is crucial. For example, an educational method that is currently popular in primary schools in the Netherlands encourages teachers and pupils to use gestures to facilitate the coupling of segments and graphemes in reading development. While it might well be the case that gestures can be helpful in this context, the types of gestures used range from iconic to metaphoric and even enactment gestures, which might influence their efficacy. In learning more about how gesture and phoneme complexity influence the efficacy of gestures in the context of L2 phoneme acquisition, we have made a start in discovering just how handy gestures can be.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion 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 Ethics Assessment Committee Humanities, Radboud University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Both authors contributed equally to the conception and design of the study, the revision of the manuscript, and read and approved the submitted version. LM performed the statistical analyses and wrote the first draft of the sections "Method" and "Results." MH wrote the first draft of the sections "Introduction" and "Discussion."

FUNDING

Open access publication fees are covered by Radboud University.

ACKNOWLEDGMENTS

Many thanks to Louis ten Bosch, Susanne Brouwer, Simon Claassen, Núria Domínguez, Lisette van der Heijden, Judith Peters, Nick Theelen, and Aurora Troncoso for their help in creating the material, data collection, and data analysis.

REFERENCES

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

Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer (Version 6.0.49) [Computer Program]*. Available online at: http://www.praat.org/ (accessed March 1, 2020).

Bohn, O.-S., and Munro, M. J. (eds). (2007). *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Q.* 22, 593–606. doi: 10.2307/3587258

Caspers, J., and Horłoza, K. (2012). Intelligibility of non-natively produced Dutch words: interaction between segmental and suprasegmental errors. *Phonetica* 69, 94–107. doi: 10.1159/000342622

Christensen, R. H. B. (2019). *Ordinal—Regression Models for Ordinal Data (Version R Package Version 2019.12-10.)*. Available online at: https://CRAN.R-project.org/package=ordinal (accessed December 10, 2019).

Collins, B., and Mees, I. (2003). *The Phonetics of English and Dutch, Revised Edition*. Leiden, NY: Brill.

Derwing, T. (2003). What do ESL students say about their accents? *Can. Mod. Lang. Rev.* 59, 547–567. doi: 10.3138/cmlr.59.4.547

Derwing, T., and Munro, M. (1997). Accent, intelligibility, and comprehensibility: evidence from four L1s. *Stud. Second Lang. Acquis.* 20, 1–16. doi: 10.1017/s0272263197001010

Derwing, T., and Munro, M. (2009). Comprehensibility as a factor in listener interaction preferences: implications for the workplace. *Can. Mod. Lang. Rev.* 66, 181–202. doi: 10.3138/cmlr.66.2.181

Escudero, P., Simon, E., and Mulak, K. E. (2014). Learning words in a new language: orthography doesn't always help. *Biling. Lang. Cogn.* 17, 384–395. doi: 10.1017/s1366728913000436

Gluhareva, D., and Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Lang. Teach. Res.* 21, 609–631. doi: 10.1177/1362168816651463

Goldin-Meadow, S. (2005). *Hearing Gesture: How Our Hands Help us Think*. Cambridge, MA: The Belknap Press.

Goldin-Meadow, S., Kim, S., and Singer, M. (1999). What the teacher's hands tell the student's mind about math. *J. Educ. Psychol.* 91, 720–730. doi: 10.1037/0022-0663.91.4.720

Graziano, M., and Gullberg, M. (2018). When speech stops, gesture stops: evidence from developmental and crosslinguistic comparisons. *Front. Psychol.* 9:879. doi: 10.3389/fpsyg.2018.00879

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (hommage a Adam Kendon). *Int. Rev. Appl. Linguist.* 44, 103–124.

Gussenhoven, C. H. M., and Broeders, A. P. A. (1997). *English Pronunciation for Student Teachers*. Winschoterdiep: Noordhoff Uitgevers.

Hannah, B., Wang, Y., Jongman, A., and Sereno, J. A. (2017). Cross-modal association between auditory and visual-spatial information in Mandarin tone perception. *J. Acoust. Soc. Am.* 140, 3225–3225. doi: 10.1121/1.4970187

Hanulíková, A., and Weber, A. (2012). Sink positive: linguistic experience with th substitutions influences nonnative word recognition. *Atten. Percept. Psychophys.* 74, 613–629. doi: 10.3758/s13414-011-0259-7

Hardison, D. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Appl. Psycholinguist.* 24, 495–522. doi: 10.1017/s0142716403000250

Hayes-Harb, R., and Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Lang. Res.* 24, 5–33. doi: 10.1177/0267658307082980

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Commun.* 47, 360–378. doi: 10.1016/j.specom.2005.04.007

Hirata, Y., Kelly, S., Huang, J., and Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *J. Speech Lang. Hear. Res.* 57, 2090–2101. doi: 10.1044/2014_jslhr-s-14-0049

Hoetjes, M., Van Maastricht, L., and Van der Heijden, L. (2019a). "Gestural training benefits L2 phoneme acquisition: findings from a production and perception perspective," in *Proceedings of the 6th Gesture and Speech in Interaction (GESPIN)*, ed. A. Grimminger (Paderborn: Universitatsbibliothek Paderborn), 50–55.

Hoetjes, M., van Maastricht, L., and Van der Heijden, L. (2019b). "Multimodal training can facilitate L2 phoneme acquisition," in *Paper Presented at the EuroSLA*, Lund.

Izura, C., Cuetos, F., and Brysbaert, M. (2014). Lextale-Esp: a test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica* 35, 49–66.

Kelly, S., Bailey, A., and Hirata, Y. (2017). Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts. *Collabra Psychol.* 3:7. doi: 10.1525/collabra.76

Kelly, S., Hirata, Y., Manansala, M., and Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Front. Psychol.* 5:673. doi: 10.3389/fpsyg.2014.00673

Kelly, S., and Lee, A. (2012). When actions speak too much louder than words: hand gestures disrupt word learning when phonetic demands are high. *Lang. Cogn. Process.* 27, 793–807. doi: 10.1080/01690965.2011.581125

Kelly, S., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Lang. Cogn. Process.* 24, 313–334. doi: 10.1080/01690960802365567

Kendon, A. (2004). *Gesture. Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kooij, J., and Van Oostendorp, M. (2003). *Fonologie: Uitnodiging tot de Klankleer van Het Nederlands*. Amsterdam: Amsterdam University 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

Li, P., Baills, F., and Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts. *Stud. Second Lang. Acquis.* (in press). doi: 10.1017/S0272263120000054

Macedonia, M., Mueller, K., and Friederici, A. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Hum. Brain Mapp.* 32, 982–998. doi: 10.1002/hbm.21084

Madan, C. R., and Singhal, A. (2012). Using actions to enhance memory: effects of enactment, gestures, and exercise on human memory. *Front. Psychol.* 3:507. doi: 10.3389/fpsyg.2012.00507

McNeill, D. (1992). *Hand and Mind. What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

Morett, L., and Chang, L. Y. (2015). Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Lang. Cogn. Neurosci.* 30, 347–353. doi: 10.1080/23273798.2014.923105

Munro, M. J., and Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 49, 285–310. doi: 10.1111/0023-8333.49.s1.8

Qualtrics (2020). *[Computer Software]*. Available online at: http://www.qualtrics.com/ (accessed March 1, 2020).

RCoreTeam (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.

Seyfeddinipur, M. (2006). *Disfluency: Interrupting Speech and Gesture*. Nijmegen: Radboud University Nijmegen.

Smotrova, T. (2017). Making pronunciation visible: gesture in teaching pronunciation. *TESOL Q.* 51, 59–89. doi: 10.1002/tesq.276

Suzukida, Y., and Saito, K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Lang. Teach. Res.* (in press). doi: 10.1177/1362168819858246

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture* 8, 219–235. doi: 10.1075/gest.8.2.06tel

Timmis, I. (2002). Native-speaker norms and International English: a classroom view. *ELT J.* 56, 240–249. doi: 10.1093/elt/56.3.240

Van den Doel, R. (2006). *How Friendly are the Natives? An Evaluation of Native-Speaker Judgements of Foreign-Accented British and American English*. Amsterdam: Netherlands Graduate School of Linguistics.

Van Maastricht, L., Hoetjes, M., and van der Heijden, L. (2019). "Multimodal training facilitates L2 phoneme acquisition: an acoustic analysis of Dutch learners' segment production in Spanish," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, eds S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Canberra: Australasian Speech Science and Technology Association Inc), 3528–3532.

Van Maastricht, L., Krahmer, E., and Swerts, M. (2016). Native speaker perceptions of (non-)native prominence patterns: effects of deviance in pitch accent distribution on accentedness, comprehensibility, intelligibility, and nativeness. *Speech Commun.* 83, 21–33. doi: 10.1016/j.specom.2016.07.008

Van Maastricht, L., Zee, T., Krahmer, E., and Swerts, M. (2020). The interplay of prosodic cues in the L2: how intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Commun.* (in press). doi: 10.1016/j.specom.2020.04.003

Van Nispen, K., van de Sandt-Koenderman, W. M. E., Mol, L., and Krahmer, E. (2016). Pantomime production by people with aphasia: what are influencing factors? *J. Speech Lang. Hear. Res.* 59, 745–758. doi: 10.1044/2015_JSLHR-L-15-0166

Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: an overview. *Speech Commun.* 57, 209–232. doi: 10.1016/j.specom.2013.09.008

Yeo, A., Wagner Cook, S., Nathan, M. J., Popescu, V., and Alibali, M. (2018). "Instructor gesture improves encoding of mathematical representation," in *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, eds T. T. Rogers, M. Rau, X. Zhu, and C. W. Kalish (Austin, TX: Cognitive Science Society), 2723–2728.

Zhang, Y., Baills, F., and Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: evidence from training Chinese adolescents with French words. *Lang. Teach. Res.* 24, 666–689. doi: 10.1177/1362168818806531

Zhen, A., Van Hedger, S., Heald, S., Goldin-Meadow, S., and Tian, X. (2019). Manual directional gestures facilitate cross-modal perceptual learning. *Cognition* 187, 178–187. doi: 10.1016/j.cognition.2019.03.004

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

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 27 November 2020
doi: 10.3389/fpsyg.2020.575952

Pantomime (Not Silent Gesture) in Multimodal Communication: Evidence From Children's Narratives

Paula Marentette[1]*, *Reyhan Furman*[2], *Marcus E. Suvanto*[3] *and Elena Nicoladis*[4]

[1]Augustana Campus, University of Alberta, Camrose, AB, Canada, [2]School of Psychology, University of Central Lancashire, Preston, United Kingdom, [3]Center for Studies in Behavioral Neuroscience, Concordia University, Montréal, QC, Canada, [4]Department of Psychology, University of Alberta, Edmonton, AB, Canada

Pantomime has long been considered distinct from co-speech gesture. It has therefore been argued that pantomime cannot be part of gesture-speech integration. We examine pantomime as distinct from silent gesture, focusing on non-co-speech gestures that occur in the midst of children's spoken narratives. We propose that gestures with features of pantomime are an infrequent but meaningful component of a multimodal communicative strategy. We examined spontaneous non-co-speech representational gesture production in the narratives of 30 monolingual English-speaking children between the ages of 8- and 11-years. We compared the use of co-speech and non-co-speech gestures in both autobiographical and fictional narratives and examined viewpoint and the use of non-manual articulators, as well as the length of responses and narrative quality. The use of non-co-speech gestures was associated with longer narratives of equal or higher quality than those using only co-speech gestures. Non-co-speech gestures were most likely to adopt character-viewpoint and use non-manual articulators. The present study supports a deeper understanding of the term pantomime and its multimodal use by children in the integration of speech and gesture.

Keywords: pantomime, co-speech gesture, non-co-speech gesture, multimodal communication, narrative, children, silent gesture, gesture-speech integration

OPEN ACCESS

Edited by:
Kazuki Sekine,
Waseda University, Japan

Reviewed by:
Ece Demir-Lira,
The University of Iowa, United States
Dipanjan Roy,
National Brain Research Centre (NBRC), India

Correspondence:
Paula Marentette
paula@ualberta.ca

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

Received: *24 June 2020*
Accepted: *04 November 2020*
Published: *27 November 2020*

Citation:
Marentette P, Furman R, Suvanto ME and Nicoladis E (2020) Pantomime (Not Silent Gesture) in Multimodal Communication: Evidence From Children's Narratives. Front. Psychol. 11:575952. doi: 10.3389/fpsyg.2020.575952

INTRODUCTION

Both pantomime and co-speech gesture refer to bodily movements used in communication (McNeill, 1992). However, pantomime has long been considered distinct from co-speech gesture. In this study, we examine representational gesture produced with and without speech in the narratives of 8–11-year-old children. We use these data to question whether there are distributional differences between spontaneously produced co-speech and non-co-speech gestures. In this paper, we argue for a distinction between two types of non-co-speech gesture: (a) silent gesture, which arises from tasks requiring communication without speech, and (b) pantomime, which, like co-speech gesture, forms a natural part of multimodal communication. In this paper, we use the term non-co-speech gesture to include all gestures produced without simultaneous speech. The terms pantomime and non-co-speech are used as they are employed by researchers when reviewing the literature. In the discussion, we address whether or not pantomime as a term can be extended to the non-co-speech gestures of the children in the present study.

The traditional definition of pantomime is variable: the central features include the absence of speech and mimetic qualities, such as the use of the whole body, and/or the adoption of a character viewpoint to enact a character's part in a narrative (McNeill, 1992; Gullberg, 1998). Pantomime as so defined is thought to contrast with co-speech gesture, which relies on its temporal links to speech for contextually specific meaning (McNeill, 1992). For example, a speaker might move the fingertips of her flattened hand upward while saying, "The jet shot up into the air."

McNeill (1992, 2016) and Levy and McNeill (2015) excluded pantomime from the gesture-language analyses, arguing that the production of pantomime by preschool children is a pragmatic attempt to facilitate an outcome rather than part of discourse. By the age of 4 years, children begin to acquire the linguistic skill and synchrony necessary for effective gesture-speech integration. By age 6, "symbolization is all in the hands" (McNeill, 2016, p. 147). By adulthood, anything that breaks this flow, such as gesture that is not aligned with speech, is "merely slovenly and not meaningful" (McNeill, 2016, p. 10).

There is reason to believe that one type of non-co-speech gesture, increasingly called "silent gesture," differs from co-speech gesture. Silent gesture occurs when participants are tasked with describing something without speaking. Adults asked to describe motion events using their hands without speech produced segmented gesture strings with consistent ordering rather than the holistic forms linked to language that are typically observed in co-speech gesture (Goldin-Meadow et al., 1996). Bilinguals asked to describe similar motion events produced different co-speech gestures depending on the language spoken: while speaking English, participants conflated manner and path gestures more often than they did while speaking Turkish (Özçalişkan, 2016; Özçalişkan et al., 2016, 2018). Critically, monolingual speakers of both languages produced conflated forms equally often in silent gesture.

There are, however, instances of similarity between silent gesture and co-speech gesture. A striking systematicity occurs in the manual representation of agentive actions compared to descriptions of objects. In comparing these representations across silent gesture and signed languages, Brentari et al. (2015) argue that these similarities arise from shared cognitive strategies aligning modes of representation with semantic categories. In particular, signers choose specific handshapes to represent the use of a tool (an agentive action) with descriptions of the tool itself (Hwang et al., 2017). Hearing non-signers using silent gesture do not demonstrate the linguistic specificity of the signers (their handshapes are not *as* selective), but they nevertheless mark the difference between actor and object (Brentari et al., 2015; see also Ortega and Özyürek, 2019). This comparison between actor and object has been extended to co-speech gesture through the analysis of gestural viewpoint (Quinto-Pozos and Parrill, 2015). ASL signers used constructed action (a linguistically embedded form of enactment) to depict the action or emotional response of characters and classifiers to depict the size, shape, or category of an object. English-speaking non-signers marked the same distinction using character-viewpoint to mark the action or emotional response of an actor and object viewpoint to mark size and shape or movement of objects (see also Gullberg, 1998, who describes the use of character viewpoint gestures as more "mimetic" than other gestures). According to Quinto-Pozos and Parrill (2015), the similarities between signers and speakers imply that this type of representation is a cognitive universal.

These findings suggest that while non-co-speech gestures may take a quasi-linguistic structure when it occurs as silent gesture in place of language, its mode of representation using viewpoints to distinguish actions with objects vs. the objects themselves may be stable regardless of accompanying speech. It is the second representational mode that may play a specific part in the non-co-speech observed in multimodal communication. In this study, we explore this representational mode in the narratives of older children.

Although we can find no explicit research on the use of non-co-speech gesture in children, children older than 6-years do use multimodal strategies in their narratives. Colletta (2009) incorporated children's use of gesture and voice in a holistic analysis of narrative development (see also Colletta et al., 2010, 2014 for a cross-cultural analysis). Alibali et al. (2009) found that, in contrast with adults, school-aged children produced more non-redundant speech-gesture combinations, with the gesture conveying somewhat different meaning than the co-occurring speech. This result suggests that the alignment between gesture and speech takes time to develop. Demir et al. (2015) report that children's use of character-viewpoint in gesture at age 5 predicted the production of more structured spoken narratives later in their development (up to age 8). Although they discuss the presence of whole-body vs. manual-only gestures, there is no mention of whether any of these character-viewpoint gestures occurred without simultaneous speech. It is worth noting that character-viewpoint gestures were relatively rare in the Demir et al. (2015) dataset. Capirci et al. (2011) explicitly coded the use of "mime" in their analysis of representational gestures in the narratives of 4–10-year old Italian children. These gestures, accounting for between 20 and 30% of the gestures, were defined as using the whole body from a character perspective, but again, there is no indication of whether these were co-speech or not.

In this study, we examine children's narratives to determine whether the distribution of non-co-speech gesture is distinct from that observed with co-speech gestures. We examined both autobiographical and fictional narratives of 8–11-year-old children. Following McNeill, we reasoned that gesture-speech integration should be adequate by this age to render the use of non-co-speech gesture unnecessary. We further examined two types of narratives to ensure we provided opportunities for distinct character-viewpoint gestures. We thought that children might be inclined to use more character-viewpoint gestures when retelling an autobiographical narrative, as these were representations of the child's own experience.

In order to determine whether there are distributional differences in children's use of non-co-speech and co-speech gesture, we pose the following questions.

- Are there differences in narrative length and quality for responses that occur with exclusively co-speech gesture, with any instance of non-co-speech gesture, or without the use of gesture at all?
- Are there differences in the features of co-speech and non-co-speech gestures? Are non-co-speech gestures more mimetic, that is, more likely to adopt a character-viewpoint or to be embodied?
- As a minor point, which type of narrative, autobiographical or fictional, is associated with the greater production of gestures with mimetic features such as embodiment or character viewpoint? We predicted that gestures in personal narratives were more likely to be produced using character-viewpoint.

MATERIALS AND METHODS

Participants

Thirty monolingual, English-speaking children (14 female) participated in this study. The children ranged in age from 8- to 11-years old (*M* = 9.7, *SD* = 12.6 months). Participants were primarily white and middle class, reflecting the demographics of the town of recruitment. Families were recruited through local posters and Facebook postings. Consent was received from parents/guardians; children provided video-recorded verbal assent for participation in this study.

Materials and Procedures

Fictional responses were elicited using two 4-min sections from Pink Panther nonverbal cartoons: *In The Pink Of The Night* (a cartoon about a cuckoo clock that bothers the Pink Panther) and *Jet Pink* (a cartoon about the Pink Panther's unskilled attempts to fly a jet plane; DePatie and Freleng, 1969-1970). The first cartoon was watched by the child and then retold to parents who had not seen the video. This process was repeated with the second cartoon. Autobiographical narratives were elicited using eight cues (see **Supplementary Table S.1**). Questions were asked in a fixed order; participants were instructed that they could pass on questions if they did not wish to answer or could not think of a response. As a result, few children responded to all autobiographical cues. This trend was apparent in pilot testing and we, therefore, used eight autobiographical cues but only two fictional cues. Children told autobiographical narratives to the researchers, who, unlike the parents, would not be familiar with the child's experiences. We chose different listeners for the stories, as we thought it likely that children would try to tell a more complete narrative to a naïve listener.

Measures and Coding

The responses were coded for length and use of representational gestures. We removed all filled pauses (e.g., "uh," "hmm," or "um") and false starts or other repeated words (McCabe et al., 2008). Remarks that did not directly relate to the narrative, such as a response to an interruption, were also removed from the count. Words that could not be transcribed (i.e., inaudible and uninterpretable) were not included in the word count (McCabe et al., 2008).

Manual iconic gestures were identified as actions with distinct strokes (McNeill, 1992) that represented information about actions, characters, objects, or events in the narratives. Embodied gestures included the use of the torso or head. Embodied gestures and iconic gestures were mutually exclusive categories. Other gestures, including deictic gestures, conventional gestures,[1] and gestures whose representational status was uncertain, were coded but are not included in this analysis. The majority of gestures produced were iconic (71%, 859 of 1,208 gestures coded).

Each representational gesture was coded for whether or not the child was speaking or silent while the stroke was produced. Recall that all gestures occur in the context of a spoken narrative, so any cessation of speech is a temporary phenomenon in this context. Sounds produced by the children were counted as onomatopoeia rather than speech as they are context-bound and depictive, rather like verbal gestures (Clark, 2016; Sasamoto and Jackson, 2016; Dingemanse, 2018). Examples are included in **Supplementary Table S.2**.

Embodied gestures included those gestures that engaged other parts of the body such as the head, legs, or torso. *Manual gestures* were limited to those produced using the hands and arms.

Viewpoint was marked for each representational gesture (McNeill, 1992; Parrill, 2010). *Observer-viewpoint* gestures use the hands to represent an object or scene. *Character-viewpoint* gestures use the hands, and sometimes the body of the storyteller to represent the hands and/or body of character in the narrative. It is possible for signers and speakers to produce a blended perspective (Dudis, 2004; Parrill, 2009). This could mean that each hand adopted a different perspective (e.g., one hand represented the cuckoo and the other the platform on which it is sitting) or that the body enacted the character while the hands depict an observer perspective (e.g., right hand representing a wall, the body representing the Pink Panther staring at it). These were coded as *blends*, but, as they were rare (*n* = 10), were analyzed as character-viewpoint in this study.

A simplified version of Stein and Albro's (1997) story grammar was used to code narrative quality. Stein and Albro (1997) identified temporal structure, causal links, goal-driven action, and the overcoming of an obstacle as components of children's narratives that indicate increasing complexity. We coded narratives into four categories. Some responses were simply *answers* to the question, not a story at all. Responses in this category did not include temporal or causal sequences. Occasionally children included a goal or outcome;

[1] Conventional gestures that were the child's commentary on the narrative ("I do not know <palms up, open hand>") were not included. Those "reported" as something the character did were included. One example was the Pink Panther patting the cuckoo bird on the head. This gestures occurred in the cartoon. Children did generate a few reported gestures such as "I do not know <palms up, open hand>" from a character's viewpoint; these were included in the analysis.

if there were no temporal or causal sequences, this was considered an answer. The inclusion of a *sequence* of events with temporal order, and sometimes causal links, was the most basic form of narrative. These responses did not include a goal or outcome. More complex narratives contained both temporal and causal sequences as well as a *goal*, giving focus to the narrative. Finally, complete narratives, called *full stories*, contained temporal and causal structure, goals, and a specified obstacle with an attempt made to overcome it. Examples of responses in each type are found in **Supplementary Table S.3**.

Analysis

The data for both variables of word length of narrative and gesture counts were highly skewed (see **Figure 1**). As a result, the analyses reported below are non-parametric. The order of telling autobiographical or fictional narratives was counterbalanced, but this did not result in any significant differences in response length, Mann-Whitney $U = 3473.5$, $p = 0.76$, or gesture counts between groups $U = 3413.5$, $p = 0.61$. As a result, data were collapsed across order for analyses.

Fictional stories were longer and accompanied by more gesture than autobiographical responses, but individual cues did not differ from each other in length or gesture count. A Kruskal-Wallis test shows that fictional responses showed higher word counts than autobiographical responses, $H(9) = 66.18$, $p < 0.001$. Dunn's *post hoc* tests showed that fictional responses did not differ between the two cartoons, $p_{\text{bonf}} = 1.00$; autobiographical responses did not differ across specific cues, $p_{\text{bonf}} = 1.00$ (except two values at 0.59 and 0.96, which are still insignificant). A Kruskal-Wallis test shows that fictional responses showed higher gesture counts than autobiographical responses, $H(9) = 32.82$, $p < 0.001$. Dunn's *post hoc* tests showed that fictional responses did not differ between the two cartoons, $p_{\text{bonf}} = 1.00$; autobiographical responses did not differ across specific cues, $p_{\text{bonf}} = 1.00$.

Reliability

Reliability was calculated for gesture identification. All responses were independently coded by two coders (the first and third authors). We calculated reliability for gesture by clause in two passes. For the first pass, we calculated linear-weighted kappa according to the following categories occurring in each entry (an entry included a full clause; a non-clause utterance, for example, "well, uh, yeah…"; or the production of the second or third gesture in a sequence): representational gesture, other gesture, and no gesture, $\kappa_w = 0.77$ ($n = 4{,}217$ entries). In this first pass, we agreed on 750 representational gestures. An additional 280 possible representational gestures were disputed. For the second pass, we independently re-coded (without discussion) these 280 disputed gestures, agreeing on a further 109. The final dataset includes a total of 859 gestures: the original 750, plus the additional 109 later-agreed gestures. A final kappa was calculated based on the categories of representational gesture and other, $\kappa_w = 0.89$.

All viewpoint decisions were coded twice (92.7% agreement). Disagreements about the viewpoint of gestures in the final dataset were discussed, with unresolved disagreements assessed as O-VPT (a more conservative code given our hypotheses).

RESULTS

Narrative Length

Children provided a total of 170 responses to fictional and autobiographical cues and produced a total of 859 gestures across 97 responses. See **Table 1** for the length of narratives and gesture production organized by whether a narrative included (i) co-speech gesture only, (ii) at least one example of non-co-speech gesture, regardless of how many co-speech gestures were produced, or (iii) no use of gesture. Note that the gesture category of responses that included one or more non-co-speech gestures *also* includes all of the co-speech gestures

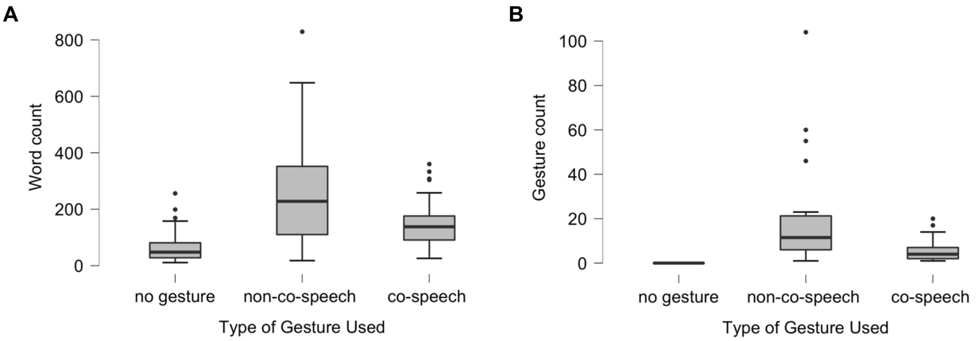


FIGURE 1 | Box plot of data counts across gesture categories including: no gesture, at least one example of non-co-speech gesture (regardless of number of co-speech gestures), or only co-speech gesture. **(A)** Reports the distribution of word count by gesture category. **(B)** Reports the distribution of gestures by gesture category. The plot is divided into quartiles: Q1 is represented by the bottom whisker, Q2 is the bottom of box to heavy line (median), Q3 is median to top of box, and Q4 is upper whisker = Q4. The dots mark outliers. The variability and outliers observed in the box plots demonstrate the non-normal distribution of data, particularly for responses that included non-co-speech gesture.

TABLE 1 | Distribution of words and gestures across narratives with differing gesture use.

	Total	Narratives with only co-speech gesture	Narratives with non-co-speech gesture(s)	Narratives with no gesture
Narrative frequency				
Total narratives	170	69	28	73
Autobiographical	116	45	11	60
Fictional	54	24	17	13
Number of children producing narratives	30	28	15[a]	25[b]
Narrative length				
Mean length in words (standard deviation)		143.1 (73.8)	267.0 (203.5)	59.9 (45.3)
Median word length		138	228	48
Word range		26–360	18–829	11–256
Gesture				
Total gesture count	859	521	338[c]	0
Mean gesture count/narrative (standard deviation)		4.9 (4.2)	18.6 (22.6)	0
Median gesture count/narrative		4	11.5	0
Gesture range		1–20	1–104	0

[a]There were no children who exclusively produced non-co-speech gestures.

[b]Two children did not gesture in any of their narratives. Many children produced one or more narratives that did not include gesture.

[c]Of the gestures produced in non-co-speech narratives, 64 were non-co-speech gestures, and the remainder were co-speech gestures.

made in that response. This is because narrative length is a property of the narrative, not of individual elements of the response (such as gesture production). Most individual children produced responses using co-speech gesture and responses using no-gesture. Half of the children in the study produced a response that included at least one non-co-speech gesture.

We tested whether gesture use was associated with response length. As response length was right skewed (a few children told very long narratives in each category, see **Figure 1A**), a non-parametric rank-based ANOVA was used. Narrative length was significantly linked with gesture category, $H(2) = 65.5$, $p < 0.001$. Dunn's *post hoc* comparisons showed that the use of either type of gesture use is associated with narratives that are significantly longer than not using gesture at all, $p < 0.001$. Narratives with non-co-speech gesture were marginally longer than stories with co-speech gesture, $p = 0.04$.

Narrative Quality

We tested whether the production of non-co-speech gesture was associated with narrative quality (see **Table 2**). Responses that included any non-co-speech gestures were most likely to be *full stories* (15/28, 53.6%), compared to responses limited to only co-speech (24/69, 34.8%) and stories with no gesture (10/73, 13.7%), χ^2 (6, $N = 170$) = 39.1, $p < 0.0001$, Cramer's $V = 0.34$, a medium effect (see **Table 2**). Stories with non-co-speech gestures were equal to or of better quality than either those with co-speech gesture or no gesture at all.

TABLE 2 | Number of narratives by story quality and gesture category.

	Story quality			
Gesture category	Answers	Sequences	Goals	Full stories
Co-speech	14	10	21	24
Non-co-speech	3	3	7	15
No gesture	43	11	9	10

Gesture categories are as follows: co-speech includes all narratives that included any co-speech gesture but no instances of non-co-speech gesture; non-co-speech includes narratives with any instance of non-so-speech gesture, regardless of how many co-speech gestures were produced; no gesture includes narratives with no instances of representational gesture.

Disentangling the relationship between narrative quality and gesture category requires consideration of the influence of narrative length (e.g., Colletta et al., 2010). This is challenging given the nominal data, non-normal distribution, and the relative rarity of non-co-speech gestures. To further explore this link, we, therefore, defined a long response as greater than or equal to the third quartile for word count in each gesture category. In **Table 3**, the counts of long responses that are *full stories* are presented as well as the counts of *full stories* that are long responses. The link between response length and narrative complexity differs by direction of effect and gesture category. In summary, for responses that included non-co-speech gesture, if the response was long, it was a *full story*, but not all *full stories* were long. The opposite trend was observed for responses that did not include gesture: Most *full stories* were long, but not all long responses were *full stories*. Responses using co-speech gesture pattern like responses with non-co-speech gesture but were somewhat less marked.

Gesture Features

Table 4 shows the distribution of co-speech and non-co-speech gestures and narrative type across articulation and viewpoint. Non-co-speech gestures (64/859, 7.4%) were less likely to occur than co-speech gestures (795/859, 92.6%). Character-viewpoint gestures (293/859, 34.1%) were less frequent than observer-viewpoint gestures (566/859, 65.9%). Embodied gestures (126/859, 14.7%) were less likely to occur than manual gestures (733/859, 85.3%).

Co-speech and non-co-speech gestures occurred proportionately across autobiographical and fictional narratives, χ^2 (1, $N = 859$) = 0.01, $p = 0.92$. Likewise, manual and embodied gestures did not differ in distribution across narrative types, χ^2 (1, $N = 859$) = 0.14, $p = 0.71$. However, distribution of viewpoint differed significantly across narrative type: In contrast to our expectations, character-viewpoint gestures constituted 58.2% of gestures in fictional stories but constituted only 26.5% of gestures in autobiographical stories, χ^2 (1, $N = 859$) = 7.85, $p = 0.006$, Cramer's $V = 0.09$, a small effect.

Gestures with mimetic features did cluster. That is, non-co-speech gestures were far more likely to be character-viewpoint and embodied (33/48, 69%), χ^2 (1, $N = 64$) = 22.71, $p < 0.0001$,

TABLE 3 | Counts (percentage) of long responses and full stories across gesture categories.

Gesture category	Long responses that are full stories	Full stories that are long responses
Co-speech (≥176 words)	15/19 (78.9%)[a]	15/24 (62.5%)
Non-co-speech (≥352 words)	7/7 (100%)	7/15 (46.7%)
No gesture (≥81 words)	7/19 (36.8%)[b]	7/10 (70.0%)

A long response was counted if the length of that story was ≥Q3 for that category.

[a] *Of four other long responses with co-speech gestures, three narratives were categorized as including goal, and one was a sequence.*

[b] *Of the 12 other long responses with no gesture, four narratives included a goal, five were sequences, and three were categorized as answers.*

Cramer's $V = 0.60$, a large effect. Gestures with this set of features are most likely to be called pantomime in the literature (e.g., Gullberg, 1998, p. 97). Co-speech gestures showed an opposite effect: they were primarily observer-viewpoint and manual (540/795, 68%), χ^2 (1, $N = 795$) = 168.65, $p < 0.0001$, Cramer's $V = 0.46$, a medium to large effect. Indeed, as can be seen in **Table 4**, there were zero non-co-speech, embodied, observer-viewpoint gestures. The 10 embodied observer-viewpoint gestures include four gestures for which there were coding disputes about viewpoint category. Recall that disputed viewpoints were coded as observer viewpoint as a more conservative decision (see Reliability section and **Supplementary Table S.2** for a description of such a gesture).

DISCUSSION

Older children did produce non-co-speech gestures as a component of their narratives. Although non-co-speech gestures were infrequent, they co-occured with other features such as character-viewpoint and embodiment. Non-co-speech gestures were associated with lengthy, high-quality stories. This examination of non-co-speech gestures challenges aspects of McNeill's position about the relationship between pantomime and gesticulation. The constellation of mimetic features observed in these narratives suggests that the use of non-co-speech gestures is an aspect of children's multimodal communication. Further, we conclude that non-co-speech gestures might be called pantomime as long as we reliably distinguish pantomime from silent gesture.

Pantomime vs. Gesticulation

McNeill's distinction between pantomime and co-speech gesture (often labeled gesticulation) arises from his exploration of the "gesture continuum." McNeill (2000) worked through the many features by which the types of gesture can be distinguished along a continuum. Relevant here is that gesticulation co-occurs with speech, pantomime does not. Pantomime is like gesticulation; however, in that, linguistics properties are absent, neither is conventionalized and they are both global in nature. Focusing on the differences between the two forms of manual activity, McNeill (2016) made three key arguments against the consideration of pantomime as part of the gesture-speech complex: that pantomime cannot orchestrate speech, that it is pragmatic, and that it occurs during a developmental stage.

We agree that non-co-speech gestures are asynchronous, and often re-enactments of an action; however, we disagree with McNeill about whether this makes these gestures pragmatic rather than symbolic. The children's production of non-co-speech gestures was integrated into a communicative act, not an effort to achieve a pragmatic outcome in their real world. We also disagree about the developmental timing of their production. Our typical and monolingual 8–11-year olds produced frequent co-speech gesture in their stories: they were not limited to non-co-speech gestures because they were unable to produce symbolic co-speech gesticulation. Much the reverse, co-speech gesture was much more frequent than non-co-speech gesture, but both types of gesture were associated with longer and more complete narratives.

All of the non-co-speech gestures produced by our participants were directly linked to the narratives they were telling. Many were linked to the surrounding speech, a few falling more closely into the category of "language-like gestures" (McNeill, 1992) as they took the place of a noun or verb in the narrative. Ladewig (2014) challenged this tendency to elevate certain forms of gesture above others. Her analysis of adults' spontaneous discourse indicated that co-speech gestures did not differ in form or function from those that occurred in language-slotted positions such as nouns or verbs. Ladewig suggested that distinctions of gesture based on their links to speech are not supported by an analysis of multimodal communication; the form and function of gestural production must be analyzed in its communicative context. Mittelberg and Evola (2014) extend this analysis with their review of the many factors, such as linguistic, discourse, and sociocultural contexts, that can influence the interpretation of the iconicity found in gestures.

We argue that the mimetic non-co-speech gestures used by children in this study were symbolic, not pragmatic in function; they were representational actions (Novack and Goldin-Meadow, 2017) serving a communicative role in the children's narratives. We turn now to an exploration of the possible role of non-co-speech gestures in multimodal communication.

Multimodal Communication

The children in this study produced gestures to support their communicative effort. It is possible that they experienced the internal cognitive benefits of gesture production (Kita et al., 2017), though that cannot be explored given our database. It is likely that non-co-speech gesture supported the external function of clearly conveying detail to the listener (de Ruiter, 2017). Mimetic gestures appear designed for the listener; as de Ruiter (2017, p. 72) suggests, the function of gesture is to "enhance the communicative signal."

Non-co-speech gestures may particularly occur when there is a notable lack of common ground (following Holler and Bavelas, 2017), that is, when the speaker is least certain of

TABLE 4 | Gesture count by articulation and viewpoint, across speech and narrative cue types.

		Co-speech gestures		Non-co-speech gestures		
Viewpoint	**Articulators**	**Autobiographical cues**	**Fictional cues**	**Autobiographical cues**	**Fictional cues**	**Total**
Character	Manual	26	136	4	11	177
	Embodied	19	64	10	23	116
Observer	Manual	159	381	3	13	556
	Embodied	2	8	0	0	10
Total		206	589	17	47	859

the recipient being able to make sense of the narrative thread. Feldman (2005) argued that mimesis is a performative act that requires interpretation in its context. Although we expected that autobiographical stories, due to their familiarity, would lead to the most character-viewpoint gestures (and given the tight link to possibly more non-co-speech gestures), this was precisely the wrong expectation. The fictional Pink Panther cartoons with their outlandish acts and unexpected turns were associated with more character-viewpoint gestures. Given the richness that is inherent in these mimetic gestures, it is possible that these were chosen because they convey details about unexpected or atypical events. For example, several children used embodied gestures (some non-co-speech) to convey the unusual turn of events when the Pink Panther burns his tail in the jet exhaust and taps the burnt end off as if it were a cigarette. That is, the use of character-viewpoint, including non-co-speech gestures, is a multimodal approach that supports effective communication.

In addition, the production of character-viewpoint gestures could lead to longer and more complex narratives. Recalling an event from an "own eyes" perspective is associated with vividness and increased details in memories of the event (Akhtar et al., 2017; St. Jacques, 2019). Perhaps a child's production of character-viewpoint gestures enhances the effects of "own eyes" recall. This, in turn, may bring to mind details of the event, leading to longer and more complex narratives. This provides a possible explanation for why non-co-speech gestures, by definition vivid, were associated with detailed and complex narratives in the present study: perhaps the use of character viewpoint had the cognitive effect of supporting memory.

The imagistic information encoded in non-co-speech gestures arises directly from the communicative goal of the speaker. Indeed, given the correlation between response length, narrative quality, and the production of non-co-speech gestures found in the present study, we argue that children use this form to support the communicative act in which they were engaged: telling "a good story." The goal of telling a "good story" may itself be enhanced through cognitive benefits of adopting a character viewpoint perspective. Categorizing all non-co-speech gesture as distinct from co-speech gesture limits our understanding of gesture-speech integration, particularly as pantomime is thought to be more common in children than in adults. We turn now to the problem of defining pantomime.

Defining Pantomime

The term pantomime incorporates many possible interpretations. It has recently been clarified by the introduction into the literature of the term "silent gesture." Silent gesture is not a typical mode of communication: It is a task assigned in the laboratory or drama studio or by necessity in particular contexts. In this study, though a few non-co-speech gestures lasted for several seconds, children did not spontaneously tell stories without recourse to speech (though many children told narratives without recourse to gesture).

Żywiczyński et al. (2018) proposed that pantomime be defined as a "communication mode that is mimetic; non-conventional and motivated; multimodal (primarily visual); improvised; using the whole body rather than exclusively manual; holistic; communicatively complex and self-sufficient; semantically complex; displaced, open-ended and universal." Żywiczyński et al. (2018) argue that this definition would exclude silent gesture (most of which are not whole body), but it may also fail to include most of the non-co-speech gestures produced by the children in this study (most are whole-body, and most are not self-sufficient). The definition proposed by Żywiczyński et al. (2018) is targeted to the question of language evolution. It would be ideal for researchers in both the language evolution community and the gesture community to embrace common definitions of terms. That will take further work and discussion.

In this paper, we seek a term to describe non-co-speech gesture that demonstrates evidence of gesture-speech integration. These are explicitly excluded from McNeill's definition of gesticulation. It is uncertain whether his use of pantomime includes the types of gesture described here. In discussing communicative dynamism, McNeill (2016) argues that what is valuable about a gesture is its ability to contribute less predictable meaning to the communicative act. From this perspective, it seems the most mimetic elements reported in this study *should* be included, as they are highly unpredictable. The meaning of the phrase "and he went <gesture>" is not interpretable without the gesture. That it is language-linked, by completing a verb slot, does not render the information less materialized or more predictable. In addition, these gestures are co-expressive, particularly if we follow the definition of the growth point as a "minimal psychological unit." In the end, we propose that the non-co-speech gestures described here do indeed orchestrate speech: on some occasions by replacing it entirely. For now, the best term to describe these appears to be pantomime.

Limitations and Future Directions

The exploration of non-co-speech gesture undertaken in this study was extensive, involving 170 narratives produced by 30 children. While this led to a reasonable 859 representational gestures, there were only 64 instances of non-co-speech gesture. Studying infrequent phenomena poses issues for typical methods of scientific analysis. The study presented here is necessarily exploratory and limited by the small sample size, both of children and, in particular, frequency of the gesture of interest. Given the results, predictive hypotheses about when children would produce non-co-speech gestures can be tested with other data sets. Further data collection should consider factors that might influence individual variation, including personality and linguistic aspects. Further qualitative analysis of the identified gestures is possible, in particular to explore their pragmatic, symbolic, and communicative functions within a linguistic system.

The explicit inclusion of non-co-speech gestures, defined as pantomime in this paper, fits into theories aiming to explain gesture-language integration. As de Ruiter (2017) points out in his rationale for the Asymmetric Redundancy–Sketch model: the link is between gesture and the communicative intention, not between gesture and local lexical items.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found at: Education and Research Archive, University of Alberta, https://doi.org/10.7939/r3-fh4t-rt03.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Alberta, Research Ethics Board Pro00041190. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PM developed the study concept and design and also drafted the first manuscript. Data collection and data coding were performed by PM and MS. RF and EN provided conceptual commentary and critical revisions. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by a grant from the Social Sciences and Humanities Research Council of Canada (IG 435-2018-004) to EN and PM, with further support from a SSHRC Institutional Grant through Office of the Vice-President, Research, of the University of Alberta, for a research assistantship to MS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.575952/full#supplementary-material

REFERENCES

Akhtar, S., Justice, L. V., Loveday, C., and Conway, M. A. (2017). Switching memory perspective. *Conscious. Cogn.* 56, 50–57. doi: 10.1016/j.concog.2017.10.006

Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., and Mainela-Arnold, E. (2009). Gesture-speech integration in narrative: are children less redundant than adults? *Gesture* 9, 290–311. doi: 10.1075/gest.9.3.02ali

Brentari, D., Di Renzo, A., Keane, J., and Volterra, V. (2015). Cognitive, cultural, and linguistic sources of a handshape distinction expressing agentivity. *Top. Cogn. Sci.* 7, 95–123. doi: 10.1111/tops.12123

Capirci, O., Cristilli, C., de Angelis, V., and Graziano, M. (2011). "Learning to use gesture in narratives: developmental trends in formal and semantic gesture competence" in *Integrating gestures*. eds. G. Stam and M. Ishino (Amsterdam: Benjamins), 187–200.

Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026

Colletta, J. -M. (2009). Comparative analysis of children's narratives at different ages: a multimodal approach. *Gesture* 9, 61–96. doi: 10.1075/gest.9.1.03col

Colletta, J. -M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., et al. (2014). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *J. Child Lang.* 42, 122–145. doi: 10.1017/S0305000913000585

Colletta, J. -M., Pellenq, C., and Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: evidence from French children and adults. *Speech Comm.* 52, 565–576. doi: 10.1016/j.specom.2010.02.009

Demir, Ö. E., Levine, S. C., and Goldin-Meadow, S. (2015). A tale of two hands: children's early gesture use in narrative production predicts later narrative structure in speech. *J. Child Lang.* 42, 662–681. doi: 10.1017/S0305000914000415

DePatie, D. H., and Freleng, F. (1969-1970). *The Pink Panther show*. Burbank, CA: DePatie-Freleng Enterprises.

de Ruiter, J. P. (2017). "The asymmetric redundancy of gesture and speech" in *Why gesture: How the hands function in speaking, thinking and communicating*. eds. R. B. Church, M. W. Alibali and S. D. Kelly (Amsterdam: Benjamins), 59–76.

Dingemanse, M. (2018). Redrawing the margins of language: lessons from research on ideophones. *Glossa* 3, 1–30. doi: 10.5334/gjgl.444

Dudis, P. G. (2004). Body partitioning and real-space blends. *Cogn. Linguist.* 15, 223–238. doi: 10.1515/cogl.2004.009

Feldman, C. F. (2005). Mimesis: where play and narrative meet. *Cogn. Dev.* 20, 503–513. doi: 10.1016/j.cogdev.2005.08.006

Goldin-Meadow, S., McNeill, D., and Singleton, J. (1996). Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychol. Rev.* 103, 34–55. doi: 10.1037/0033-295x.103.1.34

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish*. Lund, Sweden: Lund University Press.

Holler, J., and Bavelas, J. (2017). "Multi-modal communication of common ground: a review of social functions" in *Why gesture: How the hands function in speaking, thinking and communicating*. eds. R. B. Church, M. W. Alibali and S. D. Kelly (Amsterdam: Benjamins), 213–240.

Hwang, S. -O., Tomita, N., Morgan, H., Ergin, R., İlkbaşaran, D., Seegers, S., et al. (2017). Of the body and the hands: patterned iconicity for semantic categories. *Lang. Cogn.* 9, 573–602. doi: 10.1017/langcog.2016.28

Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059

Ladewig, S. (2014). "Creating multimodal utterances: the linear integration of gestures into speech" in *Body-language-communication*. *Vol. 2*. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, J. Bressem and S. Ladewig, (Berlin: De Gruyter Mouton), 1662–1677.

Levy, E., and McNeill, D. (2015). *Narrative development in young children: Gesture, imagery, and cohesion*. Cambridge: CUP.

McCabe, A., Bliss, L., Barra, G., and Bennett, M. (2008). Comparison of personal versus fictional narratives of children with language impairment. *Am. J. Speech Lang. Pathol.* 17, 194–206. doi: 10.1044/1058-0360(2008/019)

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (ed.) (2000). "Introduction" in *Language and gesture* (Cambridge: CUP).

McNeill, D. (2016). *Why we gesture*. Cambridge: CUP.

Mittelberg, I., and Evola, V. (2014). "Iconic and representational gestures" in *Body-language-communication. Vol. 2.* eds. C. Müller, A. Cienki, E. Fricke, S. Hl. Ladewig, J. Bressem and S. Ladewig (Berlin: de Gruyter Mouton), 1732–1746.

Novack, M. A., and Goldin-Meadow, S. (2017). Gesture as representational action: a paper about function. *Psychon. Bull. Rev.* 24, 652–665. doi: 10.3758/s13423-016-1145-z

Ortega, G., and Özyürek, A. (2019). Systematic mappings between semantic categories and types of iconic representations in the manual modality: a normed database of silent gesture. *Behav. Res. Methods* 52, 51–67. doi: 10.3758/s13428-019-01204-6

Özçalişkan, Ş. (2016). Do gestures follow speech in bilinguals' description of motion? *Bilingualism* 19, 644–653. doi: 10.1017/S1366728915000796

Özçalişkan, Ş., Lucero, C., and Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition* 148, 10–18. doi: 10.1016/j.cognition.2015.12.001

Özçalışkan, Ş., Lucero, C., and Goldin-Meadow, S. (2018). Blind speakers show language-specific patterns in co-speech gesture but not silent gesture. *Cogn. Sci.* 42, 1001–1014. doi: 10.1111/cogs.12502

Parrill, F. (2009). Dual viewpoint gestures. *Gesture* 9, 271–289. doi: 10.1075/gest.9.3.01par

Parrill, F. (2010). Viewpoint in speech-gesture integration: linguistic structure, discourse structure, and event structure. *Lang. Cogn. Process.* 25, 650–668. doi: 10.1080/01690960903424248

Quinto-Pozos, D., and Parrill, F. (2015). Signers and co-speech gesturers adopt similar strategies for portraying viewpoint in narratives. *Top. Cogn. Sci.* 7, 12–35. doi: 10.1111/tops.12120

Sasamoto, R., and Jackson, R. (2016). Onomatopoeia—showing-word or saying-word? Relevance theory, lexis, and the communication of impressions. *Lingua* 175–176, 36–53. doi: 10.1016/j.lingua.2015.11.003

Stein, N. L., and Albro, E. R. (1997). "Building complexity and coherence: children's use of goal-structured knowledge in telling stories" in *Narrative development: Six approaches*. ed. M. Bamberg (Mahwah, NJ: Erlbaum), 5–44.

St. Jacques, P. L. (2019). A new perspective on visual perspective in memory. *Curr. Dir. Psychol. Sci.* 28, 450–455. doi: 10.1177/0963721419850158

Żywiczyński, P., Wacewicz, S., and Sibierska, M. (2018). Defining pantomime for language evolution research. *Topoi* 37, 307–318. doi: 10.1007/s11245-016-9425-9

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 handling editor declared a past co-authorship with one of the authors RF.

BRIEF RESEARCH REPORT
published: 10 December 2020
doi: 10.3389/fpsyg.2020.587129

Gesture Influences Resolution of Ambiguous Statements of Neutral and Moral Preferences

Jennifer Hinnell[1]* *and Fey Parrill*[2]

[1] Department of English Language and Literatures, The University of British Columbia, Vancouver, BC, Canada, [2] Department of Cognitive Science, Case Western Reserve University, Cleveland, OH, United States

OPEN ACCESS

Edited by:
Autumn Hostetter,
Kalamazoo College, United States

Reviewed by:
Alexia Galati,
University of North Carolina
at Charlotte, United States
Kensy Cooperrider,
University of Chicago, United States
Leanne Beaudoin-Ryan,
Erikson Institute, United States

***Correspondence:**
Jennifer Hinnell
jennifer.hinnell@ubc.ca

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

Received: 24 July 2020
Accepted: 16 November 2020
Published: 10 December 2020

Citation:
Hinnell J and Parrill F (2020)
Gesture Influences Resolution
of Ambiguous Statements of Neutral
and Moral Preferences.
Front. Psychol. 11:587129.
doi: 10.3389/fpsyg.2020.587129

When faced with an ambiguous pronoun, comprehenders use both multimodal cues (e.g., gestures) and linguistic cues to identify the antecedent. While research has shown that gestures facilitate language comprehension, improve reference tracking, and influence the interpretation of ambiguous pronouns, literature on reference resolution suggests that a wide set of linguistic constraints influences the successful resolution of ambiguous pronouns and that linguistic cues are more powerful than some multimodal cues. To address the outstanding question of the importance of gesture as a cue in reference resolution relative to cues in the speech signal, we have previously investigated the comprehension of contrastive gestures that indexed abstract referents – in this case expressions of personal preference – and found that such gestures did facilitate the resolution of ambiguous statements of preference. In this study, we extend this work to investigate whether the effect of gesture on resolution is diminished when the gesture indexes a statement that is less likely to be interpreted as the correct referent. Participants watched videos in which a speaker contrasted two ideas that were either neutral (e.g., whether to take the train to a ballgame or drive) or moral (e.g., human cloning is (un)acceptable). A gesture to the left or right side co-occurred with speech expressing each position. In gesture-disambiguating trials, an ambiguous phrase (e.g., *I agree with that*, where *that* is ambiguous) was accompanied by a gesture to one side or the other. In gesture non-disambiguating trials, no third gesture occurred with the ambiguous phrase. Participants were more likely to choose the idea accompanied by gesture as the stimulus speaker's preference. We found no effect of scenario type. Regardless of whether the linguistic cue expressed a view that was morally charged or neutral, observers used gesture to understand the speaker's opinion. This finding contributes to our understanding of the strength and range of cues, both linguistic and multimodal, that listeners use to resolve ambiguous references.

Keywords: cohesive gesture, co-speech gesture, reference resolution, preference, contrast, discourse, multimodal communication, moral issues

INTRODUCTION

One only has to look around a room full of people spending time together to see that language consists of more than words on a page or a highly patterned audio signal. In face-to-face interaction, speakers are rarely still. Rather, in addition to the speech sounds normally associated with language, they also move their hands, shoulders, head, and manipulate their facial expressions in ways that are semantically and temporally aligned with their speech. Studies of language and cognition have thus moved beyond text and speech to include these movements as critical contributors to linguistic meaning-making (Kita, 2000, 2003; McClave, 2000; Müller et al., 2013, 2014; Levinson and Holler, 2014; Enfield, 2017; Feyaerts et al., 2017; Kita et al., 2017).

The manual gestures that speakers use in addition to speech to communicate their message are known as co-speech gestures. These gestures can be idiosyncratic and *ad hoc*, functioning "now in one way, now in another" (Kendon, 2004: 225) depending on the context. However, they are also characterized by a high degree of regularity in features such as the gesture form (Kendon, 2004; Müller, 2004), duration (Duncan, 2002), and timing of gesture related to speech (Kelly et al., 2010, 2015; Church et al., 2014; Hinnell, 2018). For example, the palm-up open-hand (PUOH) gesture is one example of a form that exhibits a stable form-meaning pairing across a speech community (Ladewig, 2014; Müller, 2017). The handshape and orientation of the PUOH are stable and iconically represent its meaning of presenting or giving information (with the open palm held in such a way as to potentially hold a small object). Similarly, a holding away gesture is prototypically enacted with both palms facing forward and raised vertically in front of the speaker; the form iconically represents how it is used, namely "to establish a barrier, push back, or hold back" a line of action, e.g., to reject topics of talk (Bressem and Müller, 2014, p. 1593; see also Kendon, 2004).

Importantly for the research presented here, speakers also use the space around their bodies in which they gesture – known as gesture space (McNeill, 1992; Priesters and Mittelberg, 2013) – in highly systematic ways to anchor objects, ideas, and other discourse elements. For example, when a speaker describes a past event and mentions that an object in the room was to the right of her, she will most likely indicate the object using a gesture to the right of her body. That is, speakers gesture in the space around their bodies to locate the things they are talking about, and, importantly, the locations of these objects in the gesture space reflect the locations of the objects in the real world. For example, we know that speakers gesture about concrete referents (objects, characters, locations) in locations in gesture space that are consistent with real locations they recall from pictures, videos, remembered events, etc. (So et al., 2009; Perniss and Ozyürek, 2015). In addition to assisting the speaker in tracking referents and building coherent discourse (McNeill, 2005; Gullberg, 2006), it's been suggested that this allows observers to use the spatial information contained in gesture to track referents and also increases comprehension (Gunter et al., 2015; Sekine and Kita, 2017).

The systematic use of gesture space also extends to abstract referents, e.g., ideas, emotions, and discourse elements (Parrill and Stec, 2017). A corpus study of English contrastive gestures showed that speakers regularly produce gestures to each side of space when contrasting two ideas (Hinnell, 2019). For example, when speakers use fixed expressions that contrast two abstract concepts, such as *on the one hand/on the other hand* or *better than/worse than*, they regularly produced gestures to each side of their body that reflect this contrastive setup, as shown in **Figure 1**. Finally, the role of space in expressing contrast extends to signed languages. For example, in American Sign Language, signers build a spatial map to make comparisons (Winston, 1996; Janzen, 2012). This comparative spatial mapping strategy has both a referential function and is used to structure discourse (Winston, 1996, p. 10).

In addition to these studies of how speakers produce gesture in contrastive discourse, experimental work has investigated how the use of gesture and gesture space affects a participant's language comprehension. Gestures that are used in establishing locations for and then tracking references in discourse are known as cohesive gestures (McNeill, 1992). It's been shown that when cohesive gestures co-occur with congruent speech, they facilitate language comprehension (Gunter et al., 2015) and can influence the interpretation of ambiguous pronouns (Goodrich Smith and Hudson Kam, 2012; Nappa and Arnold, 2014). The effect of gestures that locate referents in spatial locations extends even in the absence of the gesture. Sekine and Kita (2017) showed that listeners build a spatial representation of a story and that this representation remains active in subsequent discourse. In their study, participants were presented with a three-sentence discourse involving two protagonists. Video clips showed gestures locating the two protagonists on either side of the gesture space in the first two sentences. The third sentence referred to one of the protagonists, which could be inferred by a gendered pronoun but, importantly, did not co-occur with gesture. The name of the protagonists appeared on the screen and participants were asked to respond with one of two keys to indicate which protagonist was referred to. In the condition in which the name appeared on the side that was congruent with the gestures, participants performed better on the stimulus-response compatibility task. Importantly, there was no strategic advantage to the listeners to process the cohesive gestures, as the speech provided all information that was useful to the task (i.e., gender of protagonists). This finding extends previous findings (e.g., Goodrich Smith and Hudson Kam, 2012) that cohesive gestures allow listeners to build spatial story representations and demonstrates that listeners can "maintain the representations in a subsequent sentence without further gestural cues" (ibid: 94). In sum, listeners use a speaker's cohesive gestures to build spatial representations of concrete entities such as people or objects. This process occurs quickly (i.e., with each location mentioned or gestured once to establish a referent in a location and once again to refer back to it) and the representation remains active over the course of subsequent discourse.

Less is known about the effect on comprehension and reference resolution of gestures that contrast abstract ideas, rather than entities in narrative tasks as in the comprehension experiments described above. In previous work, Parrill and Hinnell (in review) found that observers use gesture to resolve an ambiguous statement of preference between two contrasting ideas in the same way they use gesture to resolve ambiguous references such as pronouns referring to concrete entities. That is, we found that when a speaker accompanies a statement of preference with a gesture to the same side of the gesture space that the idea was originally anchored in, the listener more frequently interprets the speaker's preference to be that idea. This suggests that people use gesture to build a spatial representation and

(1) *But apparently they like him on both sides. On the one hand, he believes in climate change so they consider him a Liberal. On the other hand, he doesn't believe in abortion so they consider him a Republican…*

FIGURE 1 | Contrastive use of gesture. 2015-09-24_1700_US_KABC_The_View, 191–201. Red Hen dataset http://redhenlab.org (click here or scan QR code to view the video clip; Uhrig, 2020).

that this representation aids listeners in resolving ambiguous references in contrastive scenarios and contributes to their understanding of a speaker's preference.

The robust literature on reference resolution provides evidence that a wide set of linguistic constraints influences the successful resolution of ambiguous pronouns. Known constraints on a listener's pronoun interpretation include linguistic salience, or conceptual accessibility. An example of linguistic salience is the subject or first-mention bias, which captures the fact that speakers most often assume the first-mentioned reference to be the referent of the ambiguous pronoun (e.g., *Francis* in the ambiguous sentence pair, *Francis went shopping with Leanne. She bought shoes*) (Gernsbacher and Hargreaves, 1988; Nappa and Arnold, 2014). Focus constructions (Arnold, 1998; Cowles et al., 2007) and recent mention (Arnold, 2001, 2010) are other linguistic constraints on reference resolution (see review in Arnold et al., 2018). Models such as Bayesian models are also based on the notion of salience. Such models suggest that reference resolution is based on probability estimates that a listener calculates based on semantic knowledge (Hartshorne et al., 2015) or from their experience of how linguistics units are used, e.g., that speakers "tend to continue talking about recently mentioned entities, especially subjects" (Arnold et al., 2018: 42; see also Arnold, 2001, 2010). As the gesture literature cited above reveals, non-verbal cues also influence pronoun interpretation; however, studies have shown that linguistic cues trump non-verbal cues during pronoun interpretation, e.g., Arnold et al. (2018) provide evidence that people rely more on their prior linguistic experience (as assessed by reading experience) than on eye-gaze aligned with the referent of the pronoun.

In light of this evidence regarding both referent tracking in multimodal contexts and reference resolution more generally, in the current study we investigate the role of gesture during the interpretation of referentially ambiguous expressions to address the relative importance of gesture as a cue in reference resolution relative to cues in the speech signal. We go beyond current literature, which has examined how gesture and gesture space are used to track concrete information (such as two entities in narrative space), to investigate the tracking of contrastive abstract information (such as pairs of moral statements). We assess whether the effect of gesture on resolving ambiguous statements is diminished when the gesture indexes a statement in speech that is less likely to be interpreted as the correct referent (e.g., a morally reprehensible position).

In this study, participants were presented with video scenarios in which the stimulus speaker contrasted two ideas. The stimulus speaker made a gesture to the left or right side that co-occurred with speech expressing each idea. Scenarios were either neutral (e.g., whether to take the train to a ball game or drive) or moral (i.e., likely to evoke strong feelings, as in human cloning is acceptable). We created two trials in which gestures were varied in the following way: in gesture-disambiguating trials, an ambiguous phrase (e.g., *I agree with that*, where *that* could refer to either previously expressed idea[1]) was accompanied by a gesture to one side or the other; in gesture non-disambiguating trials, no third gesture occurred with the ambiguous phrase. Participants were asked to identify the stimulus speaker's preference and were also asked to record their own personal preference. We explore whether participants are more likely to choose the idea accompanied by gesture as the stimulus speaker's preference (as found in earlier work), and whether this pattern changes as a function of scenario type (i.e., whether the items being contrasted were neutral in nature or involved questions of morality). We compare moral vs. neutral statements to assess whether one's own belief or that of the speaker can compete with, and potentially override, a contrastive statement of preference that is reinforced by gesture. Participants are more likely to have strong views about moral statements than about neutral statements.

This approach of considering the effect of a participant's own views on their resolution of ambiguous preference statements also aligns with an interactional approach that is gaining prominence in cognitive linguistics that considers meaning as a

[1] Other preference statements used personal pronouns, e.g., *Shelley was saying if we can clone humans, we can fix genetic disorders and end suffering. Alicia was saying there's never a good reason to go down that path. It's tough to say, but I guess I agree with her*. Here, the third person pronoun *her* could refer to either of the two underlined referents.

coordinated process between interlocutors (Clark, 1996; Du Bois, 2007; Mondada, 2013; Brône et al., 2017; Feyaerts et al., 2017). We therefore explore to what extent the participant's preference impacts the role of a co-occurring gesture on a preference statement in a contrastive scenario.

In line with literature on the role of gesture in expressing contrast and resolving ambiguous references, we hypothesized that in situations where a gesture co-occurs with one element of the contrast and then re-occurs in that place with the expression of the speaker's preference, participants would be more likely to assess this element as the speaker's preference in the scenario. Furthermore, in assessing the impact of a participants' moral views on this effect, we hypothesized that in cases where the participant disagreed strongly with the morally unacceptable position (e.g., slavery construed positively), this effect of the speaker's gesture would decrease. That is, the participant's own views would interact with the confirming effect of the gesture on how the participant assessed the speaker's preference.

The findings contribute to an understanding of the degree to which factors beyond linguistic constraints play a role in reference resolution. As cited above, Arnold et al. (2018) found gaze played less of a role than linguistic constraints in reference resolution. Here, we explore whether gesture is a powerful enough cue to resist countervailing information such as a morally abhorrent position. As such, the study contributes to our understanding of the range of cues, both linguistic and multimodal, that people recruit to resolve ambiguous references.

MATERIALS AND METHODS

Design and Predictions

We carried out a within-participants study examining the impact of two factors, scenario type (neutral, moral) and gesture trial type (gesture disambiguating, or GD; gesture non-disambiguating, or GND), on the frequency of choosing the first element of the contrast (e.g., statement A, if the contrast was A *but* B) for stimulus speaker preference (see **Figure 2** below). For the moral scenario type, the A statement always expressed the morally unacceptable option. In our earlier study (Parrill and Hinnell, in review), speakers showed a clear bias to choosing the last mentioned referent as the speaker's preference in a pair of concessive statements. Thus, the A statement was less likely to be predicted as speaker's preference. Furthermore, we operated on the assumption that having the A statement express the morally unacceptable position rendered the referent more predictable, as the B statement was more likely to represent the speaker's intended position. We also predicted that participants would be more likely to choose the A statement for stimulus speaker preference when the speaker makes a disambiguating gesture, i.e., for GD trials as compared to GND trials. If it is the case that gesture plays less of a role when the majority of participants disagree with the position expressed in the A statement, then we would expect the frequency of those choosing A to decrease for moral scenarios as compared to neutral scenarios within the GD trials.

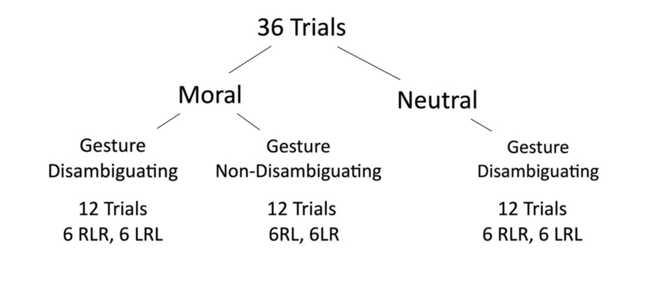


FIGURE 2 | Experimental design.

Materials

We created 36 scenarios, each containing the following elements:

(1) An attitude about a topic (A statement).
(2) The concessive "but."
(3) A differing attitude about the topic (B statement).
(4) A hedge indicating uncertainty.
(5) An ambiguous statement indicating a preference for either the A or B statement[2].

For example, "My little brother's not on Facebook because he thinks it's a waste of time" (A statement), "but" (concessive) "my other brother says he can't do job networking without it" (B statement). "I can see what they're getting at, but" (hedge) "I think he's right" (preference statement). The preference statement is ambiguous because "he" could refer to either "little brother" or "other brother."

We created two types of scenarios, neutral and moral. For neutral scenarios, we used previous research (Parrill and Hinnell, in review) as a starting point. We selected twelve scenarios for which participants in the previous study chose the A and B statements at about equal rates when asked about their own personal preference. Returning to the example given above, about half the participants in our previous study thought Facebook is a waste of time and about half thought Facebook is useful. We created 24 moral scenarios based on topics selected from Gallup's annual Values and Beliefs poll (Jones, 2017) and a study of divisive social issues (Simons and Green, 2018). Topics were included if at least 70% of participants in these sources considered one position related to the topic morally unacceptable. We then created scenarios about these topics. Moral scenarios always had the following form: An A statement that expressed the morally *unacceptable* position, the concessive "but," a B statement that expressed the morally *acceptable* position, a hedge, and an ambiguous preference statement. For example, 86% of participants in the Gallup study considered human cloning morally unacceptable, so human cloning was included as a topic. An example scenario is: "Shelley was saying if we can clone

[2] While maintaining a fairly constrained template for the preference statement (see full stimuli in **Supplementary Material**), 2 of the 3 preference statements we used included a second hedge within the preference statement, e.g., HEDGE + *I guess I agree with her*. Given the results of recent corpus studies (Hinnell, 2019, 2020) and current research on "stance stacking," which has shown that most frequently, highly stanced elements co-occur with each other (i.e., are "stacked," Dancygier, 2012), we incorporated this type of more natural speech in our stimuli.

humans, we can fix genetic disorders and end suffering" (A statement, morally unacceptable position), "but" (concessive) "Alicia was saying there's never a good reason to go down that path" (B statement, morally acceptable position). "It's tough to say, but" (hedge) "I guess I agree with her" (preference statement). The preference statement is ambiguous because "her" could refer to either Shelly or Alicia.

We first recorded audio for each scenario. The first author read each scenario as naturally as possible. For the recording of video, a research assistant was instructed to sit in a comfortable posture and to perform (speak and gesture) several scenarios as naturally as possible. Scenarios were performed in two different ways: a gesture-disambiguating version (GD) and a gesture non-disambiguating version (GND).

Both the GD and GND versions of the video featured palm-up open-hand gestures (see **Figure 3**). These were performed with the A and B statements[3]. The research assistant performed versions with the left hand first and with the right hand first. For the GND version, the speaker sat still and did not gesture during the preference statement. For the GD version, the speaker performed a final palm-up open-hand gesture with the preference statement. The final gesture always occurred in the location where the A statement gesture had been performed. For example, if the first gesture was on the left, the final gesture would be performed on the left as well.

We created four types of videos: (1) right hand first, left hand second, final gesture with right hand, (2) left hand first, right hand second, final gesture with left hand, (3) right hand first, left hand second, no third gesture, and (4) left hand first, right hand second, no third gesture. Using Final Cut Pro, we matched audio clips to these four different videos to create two stimulus lists. We used stimulus lists to minimize the chances that specific properties of the scenarios would impact our results.

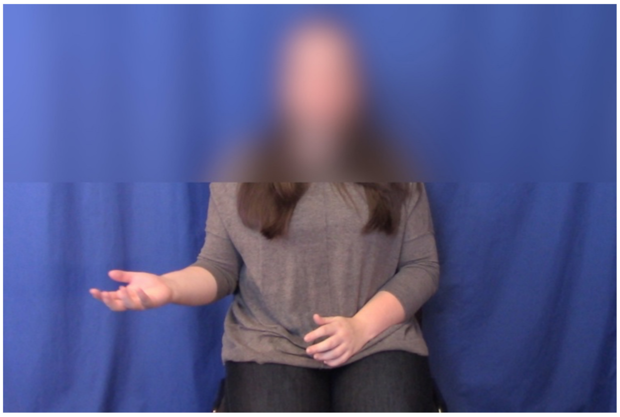


FIGURE 3 | Example stimulus.

Scenarios were assigned to GD and GND videos to create 12 moral GND trials and 12 moral GD trials. Because our previous study indicated that we could not include more than 36 trials without participants becoming fatigued, we created only GD neutral trials. This design was selected to maximize our ability to compare *moral* scenarios across gesture disambiguating and non-disambiguating trials, without the study lasting so long that participants would not be able to attend to the stimuli. We elected to use a smaller number of neutral scenarios, and to use only GD trials for our neutral scenarios, because our previous work indicated that without gesture, participants will choose the B statement as the speaker's preference at a rate above chance (about 70%). Moral scenarios were counterbalanced across stimulus lists so that each occurred with both GD and GND videos.

When adding audio to video, we aligned specific auditory and gestural features. Gesture strokes (the effortful, meaningful portion of a gesture: McNeill, 1992) were aligned with the subject noun phrases (e.g., "little brother," "other brother"). The stroke of the final gesture for GD stimuli was aligned with the ambiguous noun phrase (e.g., "he's"). We used Final Cut Pro to blur the speaker's face and upper shoulders so that mouth movements did not reveal the fact that audio and video had been edited, as shown in **Figure 3**. We also did this masking so that facial expressions and head movements would not affect participants' judgments. There was some variation in intonational contours and in the research assistant's posture across different videos. This was desirable, as it made the scenarios feel more natural.

In summary, the outcome of the editing was to create two versions of each scenario, with scenarios randomly paired to GD and GND videos for the moral scenarios, and always paired with GD videos for the neutral scenarios. Within moral and neutral categories, scenarios were randomly paired with videos in which the right versus left hand was used first. Audio and video were carefully aligned to preserve the systematicity of auditory and gestural cues. Participants were presented with both neutral and moral scenarios and both GD and GDN (a within-participants study). Trials were presented in random order.

Procedure

After an informed consent/instruction screen, participants were presented with a scenario. After viewing each scenario, participants responded to a question asking for their judgment about the stimulus speaker's preference. The exact question was matched to the preference statement, so that, for example, a preference statement ending with "I think he's right" would be followed by a question asking "who does the speaker think is right?" Participants chose between options matched to the scenario, such as "Facebook is a waste of time" and "Facebook is needed for networking." Options were presented horizontally, and their locations were random (thus, the option appearing to the left was random for each trial so that the choice options didn't necessarily match the spatial location of the A and B statements). Second, participants responded to the question "What is your personal opinion/preference?" and were presented with the same options as in the previous variable (e.g., "Facebook is a waste of time," "Facebook is needed for networking"). As with the previous

[3] We refer to the two statements throughout the paper as A and B statements to be consistent between moral and neutral trials (i.e., neutral trials have no morally unacceptable or acceptable positions). In cases where we discuss moral scenarios only, we will use morally unacceptable (A) and morally acceptable (B) for ease of comprehension.

response, the location of options was randomized with respect to location. Responses to these two questions serve as our dependent variables and will be referred to as "stimulus speaker preference" and "participant preference." After the last scenario, participants answered demographic questions about gender (male, female, other), race, age, fluency in a second language, political ideology ("do you identify as more progressive/more conservative"), and participants were asked "what do you think this study was about?" (open entry).

Participants

Eighty participants were recruited using the online data collection platform Amazon Mechanical Turk. Mechanical Turk data have been shown to be comparable to data collected in academic research studies, but the Mechanical Turk population is more diverse in age, education, and race/ethnicity than most typical university research populations (Burhmester et al., 2011). Participants were required to be within the United States to take part and were compensated with $3.50. The study took about half an hour for participants to complete.

RESULTS

Data have been uploaded to Open Science Framework and can be found here: https://osf.io/t3sbx/. Data were examined to ensure no participants completed the study too quickly to have done the task correctly. One participant was removed from the List 1 data for this reason. Demographic details (age, gender, race, political affiliation) are presented in the **Supplementary Appendix 1**. When asked about the topic of the study, only six of the 80 participants said that the study was about gesture or body language (these six were not removed from the analyses). The majority of participants said that the study was about things like persuasion, decision making, or opinions. Data were analyzed using R version 4.0.0 (R Core Team, 2020). Utility packages used for data manipulation, cleaning, and analysis include dplyr (Wickham et al., 2020), tidyr (Wickham and Henry, 2020), psych (Revelle, 2019), car (Fox and Weisberg, 2019), lawstat (Hui et al., 2008), and DescTools (Signorell et al., 2020).

We present the proportion of participants who chose the A statement as their own personal preference/opinion for each scenario in the **Supplementary Appendix 2** along with the scenario texts. In general, the scenarios patterned as expected (neutral scenarios around 50%, moral scenarios below 30%). There were some exceptions (to which we will return in the discussion), but this is not problematic for our predictions. The majority of the scenarios behaved as expected and we examine frequency data.

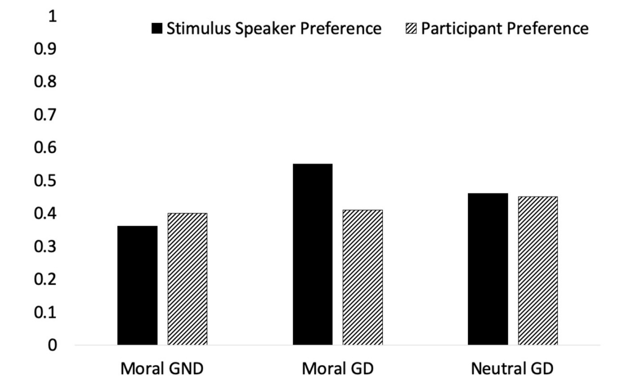


FIGURE 4 | Proportion of participants who chose the A statement by condition and scenario (no error bars shown because the figure shows overall proportions).

TABLE 1 | Responses by scenario type and trial type (proportions in parentheses).

Scenario type	Trial type	Response*	Stimulus speaker preference	Participant preference
Neutral	GD	A	436 (0.46)	423 (0.45)
		B	512 (0.54)	525 (0.55)
Moral	GND	A	345 (0.36)	376 (0.40)
		B	603 (0.60)	572 (0.60)
	GD	A	524 (0.55)	385 (0.41)
		B	424 (0.45)	563 (0.59)

**For moral scenario types, A was the morally unacceptable response, B was the morally acceptable response.*

Because our data are categorical and do not meet the assumptions required for parametric tests (they are non-normal, non-interval, and we do not have homogeneity of variance), we used chi-square analyses to answer our research questions. These analyses mean that we will not examine some possible relationships (how different scenarios might pattern, variability contributed by participants, etc.). However, these analyses were preferable to logistic regression as they require fewer assumptions about the data and are simpler to interpret.

Figure 4 shows the proportion of participants who chose the A statement for stimulus speaker preference. **Table 1** shows an overall picture of the data both as frequencies and proportions according to scenario type and trial type. The key comparison is between the proportion of participants choosing the A statement for stimulus speaker preference. This proportion is higher for both types of GD trials (46% and 55%) compared to GND trials (36%).

Table 2 presents responses according to what the participant selected for both dependent variables, by trial type and list. That is, 117 participants chose A for both stimulus speaker preference and participant preference for the Moral GND trials for list 1. While this presentation of the data is not as easy to relate to the research questions as **Table 1**, the contingency tables created allow us to use a variant of the chi-square test that accounts for multiple dimensions, called the Cochran-Mantel-Haenszel chi-square test. This test creates a common odds ratio (OR) across multiple contingency tables, which allows researchers to avoid Simpson's paradox (Simpson, 1951), wherein patterns that appear when comparing one subset of the data disappear when comparing another subset. ORs are a conditional estimate of the extent to which a treatment impacts an outcome (e.g., the odds of choosing the A statement for stimulus speaker preference

TABLE 2 | Stimulus speaker and participant preference by trial type and list, with odds ratios.

	Stimulus speaker preference*	Participant preference		Odds ratio**
		A	B	
Moral, GND, List 1	A	117	72	4.52
	B	77	214	
Moral, GND, List 2	A	111	58	6.17
	B	80	258	
Moral, GD, List 1	A	145	126	3.66
	B	50	159	
Moral, GD, List 2	A	125	128	2.52
	B	60	155	
Neutral, GD, List 1	A	141	118	2.63
	B	69	152	
Neutral, GD, List 2	A	136	117	2.08
	B	77	138	

*For moral scenario types, A was the morally unacceptable response, B was the morally acceptable response.

**Odds ratio presents the odds of choosing B for participant preference given person chose B for stimulus speaker preference.

given you chose A for participant preference). An OR close to 1 indicates no impact on outcome (outcome is 1 time as likely). Overall, these analyses test a null hypothesis that the choice between A and B for stimulus speaker preference is not independent of the choice for participant preference.

The CMH statistic of 228.45 (1), $p < 0.0001$ (pooled OR = 3.26) indicates a significant association between one of our variables and outcomes. This leads us to reject the null hypothesis that the dimensions are independent. We then tested the homogeneity of ORs using the Breslow-Day test, which tests the null hypothesis that the ORs are all statistically the same. R's DescTools allows the Breslow-Day test to be calculated with or without Tarone's adjustment; we opted to calculate without because we have a relatively large sample size and the need for more accurate p-values was moot. The Breslow-Day chi-square statistic [X2 (5, $N = 2883$) = 21.14, $p = 0.0008$] indicates that the ORs are not the same.

Table 2 shows the individual ORs for each by-list contingency table. In general, participants tend to choose B for both dependent variables (that is, they "pile up" in the B/B corner of the tables). The odds of this are particularly high when there is no disambiguating gesture (between 4 and 6 times as likely).

To provide some statistical information about the impact of list, we compared the two moral GD contingency tables (that is, across list 1 and list 2). Here the Breslow-Day chi-square statistic [X2 (1) = 1.73, $p = 0.19$] requires us to fail to reject the null hypothesis that the two ORs are statistically equivalent. This indicates that the association is not based on list for moral GD trials. A comparison of the two moral GND contingency tables across list also requires us to fail to reject the null hypothesis that the two ORs are the same [X2 (1) = 1.18, $p = 0.28$]. This indicates that the association is not based on list for moral GND trials. Finally, a comparison of the two neutral GD contingency tables (across lists) also requires us to fail to reject the null hypothesis that the two ORs are the same [X2 (1) = 0.75, $p = 0.39$]. This indicates that the association is not based on list for neutral GD trials. Taken together, this set of analyses indicates that the lists can be collapsed, thus we aggregated the data across lists.

To determine the impact of trial type (with a final disambiguating gesture, without a final gesture), we first compared moral GND to moral GD trials. The Breslow-Day chi-square statistic indicates that there is an association between trial type and outcome [X2 (1) = 7.56, $p = 0.006$]. For moral scenarios, participants were more likely to choose the A statement when a gesture was produced on the "A side" with the preference statement, compared to when there was no gesture.

To understand the impact of scenario type (moral, neutral), we compared moral GD trials to neutral GD trials. The Breslow-Day chi-square statistic indicates that there was no association between scenario type and outcome [X2 (1) = 1.77, $p = 0.18$]. Participants were equally likely to choose the A statement for moral GD and neutral GD trials.

Finally, we verified that gesture was used to disambiguate preference across scenario types by comparing the moral GND trials to the neutral GD trials. The Breslow-Day chi-square statistic indicates that there is an association between scenario type and outcome [X2 (1) = 17.08, $p < 0.00001$]. Participants were more likely to choose the A statement when a gesture was produced on the "A side" with the preference statement (neutral GD trials), compared to when there was no gesture (moral GND trials).

DISCUSSION

We predicted that when presented with scenarios in which the speaker produced an ambiguous expression of preference, participants would use gesture to disambiguate, if gesture was available. That is, if the speaker produced a gesture in the location where she had previously gestured when presenting a position, participants would be more likely to assume she preferred that option. This prediction was supported. Participants were more likely to choose the A statement when a gesture was produced in the "A location" during the preference statement. This replicates our previous work, showing that gesture is integrated into participants' understanding of a speaker's preference. In the context of research on cohesive gestures and reference resolution, this finding provides further evidence that gesture is recruited by the listener to resolve ambiguous references.

Beyond this, we extended our previous work by asking whether gesture as a cue in reference resolution would play less of a role when the position expressed in the A statement was an unpopular one. That is, if the speaker appeared to indicate via the location of her gesture that she was in favor of slavery, would participants be more likely to ignore her gesture and assume she preferred the more acceptable B statement position? In fact, we found no effect of scenario type. Participants were equally likely to choose the A statement when a gesture in the "A location" occurred with the preference statement regardless of whether the scenario was a moral or a neutral one. This finding suggests that gesture is a relatively strong referential cue, i.e., it can influence listeners to select an intended referent even when the referent indexes countervailing contextual information such as a morally unacceptable position.

While the presence of gesture shifted participants' assessment of the speaker's preference, participants still chose the B statement (whatever came last) between 40 and 60% of the time. In these cases, the linguistic cue appears to override the gestural cue. Even though a participant was using gesture to indicate a preference for a position that is relatively unpopular (i.e., in moral scenarios), the pattern was the same. Further studies are needed to explore whether gesture plays a more prominent role when the linguistic cue is weaker.

While Arnold et al. (2018) found that people rely more strongly on linguistic experience than on eye gaze, our findings suggest that contextual information such as a speaker's predicted preference can indeed be "trumped" by gesture in the resolution of ambiguous reference. In our study, participants relied on the gestural cue for the morally unacceptable scenarios, despite most of the participants indicating they were not explicitly aware of the gestures, or at least of gesture as a point of the study. While not necessarily at odds with the finding of Arnold et al. (2018) given that here we examine the role of gesture rather than gaze, which may be a weaker cue, our findings underscore the need for further studies that include a range of linguistic, contextual, and multimodal cues to assess their relative strengths in reference resolution contexts.

Although the majority of our scenarios patterned the way we expected them to (that is, were neutral or moral, according to the way we operationalized these concepts for this project), there were some interesting exceptions. Participants in our data were more favorable toward human cloning, high unemployment, vandalism, air pollution, and polygamy than predicted. It is important to note that our scenarios justified a particular position (e.g., human cloning is good because it can end human suffering), whereas the research we were drawing from only presented a topic and asked participants to align as pro or con. It is also worth noting that 58% of our participants identified as politically conservative and that our data were collected in June, 2020. This was a highly atypical historical moment, as the United States was experiencing record unemployment due to the COVID-19 global pandemic in addition to sustained national protests over police brutality and racial injustice. This may have had some impact on responses to human cloning (as a means of curing disease), high unemployment, vandalism (framed as an act of protest in the scenario), and polygamy (framed as sharing the burden of childcare in the scenario). There were also two neutral scenarios where participants chose the A statement at rates considerably below 50%. Again, because our analyses are frequency based, these exceptions are not problematic, but do underscore the variability in opinion that makes such research challenging.

Another limitation of the current study was in the variability of the stimuli. Some of the preference statements included a second hedge within the preference statement, e.g., *I don't know, but I guess I agree more with that*, as opposed to *hard to know, but his argument makes more sense to me.* Although this may have introduced more variability, this was done to incorporate the most natural speech possible in an experimental context; corpus studies have shown that speakers very frequently "stack" highly stanced elements such as hedges (Hinnell, 2019, 2020).

Another factor that impacted the naturalness of the stimuli was the decision to obscure the face of the speaker. This was done to remove the possibility of mouth movements revealing the fact that audio and video had been edited. Since gaze is frequently where interlocutors fixate when interacting with a speaker, this frequently used stimuli design may push the listener to pay more attention to the hands than they normally would [i.e., listeners tend not to attend to speakers gestures directly (Gullberg and Kita, 2009)]. We have attempted to mitigate the impact of this design somewhat through our debriefing process, in which we asked participants what they thought the study was about. Responses indicate that gesture was not very salient[4]. Finally, though we collected demographic data, we have not analyzed them in detail, planning instead to include them in future studies. It may be that additional patterns emerge when we examine sex, race, age, or political identification (though our measure of this was quite gross, being only a binary choice between more progressive and more conservative).

Several further questions remain. Firstly, in this study the stimulus speaker gestured only with PUOH gestures. However, the corpus studies in Hinnell (2019, 2020) suggest that speakers also regularly use other hand forms as well as other body articulators (e.g., head movements side to side) to indicate contrast, particularly when the referents are abstract. The question arises, then, whether other handshapes would affect the comprehension of contrastive gestures of preference and whether the effect is the same if the contrast is indicated in the head rather than the hands. That is, do hand form and articulator influence comprehension as well as placement in gesture space. Secondly, participants in this study were a variety of ages (mean age 36). Sekine and Kita (2015) have shown that children ∼5 years of age fail to integrate spoken discourse and cohesive use of space in gestures. We would expect that children of this age would also fail to integrate gestures of preference as explored in this study at that age, acquiring this ability before the age of 10 (in Sekine and Kita's study, 10 year-olds performed the same as adults).

In sum, in this study we explored the effect of gesture on the observer in contrastive discourse, examining in particular the effect of gesture when speakers were expressing preference about neutral vs. highly moral issues. Findings suggest that gesture disambiguates an expression of the speaker's preference for the observer. This contribution does not change even when the view being expressed is contrary to the participants' beliefs and might be seen as socially unacceptable (e.g., the suggestion that slavery had benefits). These findings extend the scope of reference resolution studies beyond concrete referents in narrative storytelling to contrastive scenarios involving abstract referents. Furthermore, as one of few reference resolution studies to evaluate the strength of gesture in light of contextual cues, it points to the need to include multimodal cues in

[4] As one reviewer pointed out, this could be because gesture was not particularly salient, however, this could also be because participants thought gesture was so central to the study it did not bear mentioning. A more structured debriefing would help us evaluate this for future studies.

reference resolution studies and underscores the importance of gesture in creating multimodal discourse.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found online at: https://osf.io/t3sbx/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Case Western Reserve University Institutional Review Board, DHHS FWA00004428 and IRB registration number 00000683. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Both authors created the experimental design and the neutral and moral scenarios used in the materials and wrote the manuscript. JH recorded the audio stimuli and prepared the manuscript for submission. FP coordinated video production/editing, prepared the surveys for data collection, collected the data, and performed the statistical analysis.

FUNDING

Publication fees were supported by the Cognitive Science Department, Case Western Reserve University.

ACKNOWLEDGMENTS

We are grateful to Matthew Pietrosanu of the University of Alberta's Training and Consulting Centre (TCC) for advice on statistical procedures and to three reviewers for their constructive feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.587129/full#supplementary-material

REFERENCES

Arnold, J. E. (1998). *Reference Form and Discourse Patterns*. Doctoral dissertation, Stanford University, Stanford, CA.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Process.* 31, 137–162. doi: 10.1207/s15326950dp3102_02

Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Lang. Linguist. Compass* 4, 187–203. doi: 10.1111/j.1749-818x.2010.00193.x

Arnold, J. E., Strangmann, I., Hwang, H., Zerkle, S., and Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *J. Mem. Lang.* 102, 41–54. doi: 10.1016/j.jml.2018.05.002

Bressem, J., and Müller, C. (2014). "The family of away gestures," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Teßendorf (Berlin: De Gruyter Mouton), 1592–1604.

Brône, G., Oben, B., Jehoul, A., Vranjes, J., and Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. *Cogn. Linguist.* 28, 449–483. doi: 10.1515/cog-2016-0119

Burhmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* 6, 3–5. doi: 10.1177/1745691610393980

Church, R. B., Kelly, S., and Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Lang. Cogn. Neurosci.* 29, 345–354. doi: 10.1080/01690965.2013.857783

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539

Cowles, H. W., Walenski, M., and Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi* 26, 3–18. doi: 10.1007/s11245-006-9004-6

Dancygier, B. (2012). "Negation, stance verbs, and intersubjectivity," in *Viewpoint in Language: A Multimodal Perspective*, eds B. Dancygier, and E. Sweetser (Cambridge, NY: Cambridge University Press), 69–93. doi: 10.1017/cbo9781139084727.006

Du Bois, J. W. (2007). "The stance triangle," in *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, ed. R. Englebretson (Amsterdam: John Benjamins), 138–182. doi: 10.1075/pbns.164.07du

Duncan, S. (2002). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture* 2, 183–206. doi: 10.1075/gest.2.2.04dun

Enfield, N. (2017). *How We Talk: The Inner Workings of Conversation*. New York, NY: Basic Books.

Feyaerts, K., Brône, G., and Oben, B. (2017). "Multimodality in interaction," in *The Cambridge Handbook of Cognitive Linguistics*, ed. B. Dancygier (Cambridge: Cambridge University Press), 135–156. doi: 10.1017/9781316339732.010

Fox, J., and Weisberg, S. (2019). *An {R} Companion to Applied Regression*, 3rd Edn. Thousand Oaks, CA: Sage.

Gernsbacher, M. A., and Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *J. Mem. Lang.* 27, 699–717. doi: 10.1016/0749-596X(88)90016-2

Goodrich Smith, W., and Hudson Kam, C. K. (2012). Knowing 'who she is' based on 'where she is': the effect of co-speech gesture on pronoun comprehension. *Lang. Cogn.* 4, 75–98. doi: 10.1515/langcog-2012-0005

Gullberg, M. (2006). Handling discourse: gestures, reference, tracking, and communication strategies in early L2. *Lang. Learn.* 56, 155–196. doi: 10.1111/j.0023-8333.2006.00344.x

Gullberg, M., and Kita, S. (2009). Attention to speech-accompanying gestures: eye movements and information uptake. *J. Nonverbal Behav.* 33, 251–277. doi: 10.1007/s10919-009-0073-2

Gunter, T. C., Weinbrenner, J. E. D., and Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Front. Psychol.* 6:80. doi: 10.3389/fpsyg.2015.00080

Hartshorne, J. K., O'Donnell, T. J., and Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Lang. Cogn. Neurosci.* 30, 716–734. doi: 10.1080/23273798.2015.1008524

Hinnell, J. (2018). The multimodal marking of aspect: the case of five periphrastic auxiliary constructions in North American English. *Cogn. Linguist.* 29, 773–806. doi: 10.1515/cog-2017-0009

Hinnell, J. (2019). The verbal-kinesic enactment of contrast in North American English. *Am. J. Semiotics* 35, 55–92. doi: 10.5840/ajs20198754

Hinnell, J. (2020). *Language in the Body: Multimodality in Grammar and Discourse*. Doctoral dissertation. University of Alberta, Edmonton, AB. doi: 10.7939/r3-1nhm-5c89

Hui, W., Gel, Y., and Gastwirth, J. (2008). lawstat: an R package for law, public policy and biostatistics. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i03

Janzen, T. (2012). "Two ways of conceptualizing space: motivating the use of static and rotated vantage point space in ASL discourse," in *Viewpoint in Language*, eds B. Dancygier, and E. Sweetser (New York, NY: Cambridge University Press), 156–176. doi: 10.1017/cbo9781139084727.012

Jones, J. M. (2017). *Americans Hold Record Liberal Views on Most Moral Issues*. Washington, DC. Available online at: https://news.gallup.com/poll/210542/americans-hold-record-liberal-views-moral-issues.aspx (accessed February 13, 2020).

Kelly, S., Healey, M., Özyürek, A., and Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychon. Bull. Rev.* 22, 517–523. doi: 10.3758/s13423-014-0681-7

Kelly, S., Ozyürek, A., and Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572

Kita, S. (2000). "How representational gestures help speaking," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 162–185. doi: 10.1017/cbo9780511620850.011

Kita, S. (2003). *Pointing: Where Language, Culture, and Cognition Meet*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. doi: 10.4324/9781410607744

Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059

Ladewig, S. H. (2014). "Recurrent gestures," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1558–1574.

Levinson, S. C., and Holler, J. (2014). The origin of human multi-modal communication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130302. doi: 10.1098/rstb.2013.0302

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *J Pragmat.* 32, 855–878. doi: 10.1016/s0378-2166(99)00079-x

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226514642.001.0001

Mondada, L. (2013). "Conversation analysis: talk & bodily resources for the organization of social interaction," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 1, eds C. Müller, E. Fricke, S. H. Ladewig, S. Tessendorf, and D. McNeill (Berlin: De Gruyter Mouton), 218–226.

Müller, C. (2004). "Forms and uses of the palm up open hand: a case of a gestural family," in *The Semantics and Pragmatics of Everyday Gestures*, eds C. Müller, and R. Posner (Berlin: Weidler), 233–256.

Müller, C. (2017). How recurrent gestures mean: conventionalized contexts-of-use and embodied motivation. *Gesture* 16, 276–303. doi: 10.1075/gest.16.2.05mul

Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., and Bressem, J. (Eds). (2014). *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2. Berlin: De Gruyter Mouton.

Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., and Teßendorf, S. (Eds). (2013). *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 1. Berlin: De Gruyter Mouton.

Nappa, R., and Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: cues to the speaker's attention and intentions affect pronoun comprehension. *Cogn. Psychol.* 70, 58–81. doi: 10.1016/j.cogpsych.2013.12.003

Parrill, F., and Hinnell, J. (in review). *Observers use Gesture to Disambiguate Contrastive Expressions of Preference*.

Parrill, F., and Stec, K. (2017). Gestures of the abstract: do speakers use space consistently and contrastively when gesturing about abstract concepts? *Pragmat. Cogn.* 24, 33–61. doi: 10.1075/pc.17006.par

Perniss, P., and Ozyürek, A. (2015). Visible cohesion: a comparison of reference tracking in sign, speech, and co-speech gesture. *Top. Cogn. Sci.* 7, 36–60. doi: 10.1111/tops.12122

Priesters, M., and Mittelberg, I. (2013). Individual differences in speakers' gesture spaces: multi-angle views from a motion-capture study. *Paper Presented at the Proceedings of TiGeR 2013*, Tilburg, NL.

Revelle, W. (2019). *psych: Procedures for Personality and Psychological Research*. Evanston, IL: Northwestern University.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Sekine, K., and Kita, S. (2015). Development of multimodal discourse comprehension: cohesive use of space by gestures. *Lang. Cogn. Neurosci.* 30, 1245–1258. doi: 10.1080/23273798.2015.1053814

Sekine, K., and Kita, S. (2017). The listener automatically uses spatial story representations from the speaker's cohesive gestures when processing subsequent sentences without gestures. *Acta Psychol.* 179, 89–95. doi: 10.1016/j.actpsy.2017.07.009

Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arachchige, C., et al. (2020). *DescTools: Tools for descriptive statistics. R package version 0.99.38*.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *J. R. Stat. Soc. Series B.* 13, 238–241. doi: 10.1111/j.2517-6161.1951.tb00088.x

Simons, J. J. P., and Green, M. C. (2018). Divisive topics as social threats. *Commun. Res.* 45, 165–187. doi: 10.1177/0093650216644025

So, W. C., Kita, S., and Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cogn. Sci.* 33, 115–125. doi: 10.1111/j.1551-6709.2008.01006.x

Uhrig, P. (2020). Multimodal research in linguistics. *Zeitschrift für Anglistik und Amerikanistik* 6, 345–349. doi: 10.1515/zaa-2020-2019

Wickham, H., François, R., Henry, L., and Müller, K. (2020). *Dplyr: A Grammar of Data Manipulation. R Package Version 0.8.5*.

Wickham, H., and Henry, L. (2020). *Tidyr: Tidy Messy Data. R Package Version 1.0.2*.

Winston, E. A. (1996). Spatial mapping in ASL discourse. *Paper Presented at the CIT 11th National Convention*. Little Rock, AR.

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.

CORRECTION
published: 05 March 2021
doi: 10.3389/fpsyg.2021.664194

Corrigendum: Gesture Influences Resolution of Ambiguous Statements of Neutral and Moral Preferences

Jennifer Hinnell[1]* *and Fey Parrill*[2]

[1] Department of English Language and Literatures, The University of British Columbia, Vancouver, BC, Canada, [2] Department of Cognitive Science, Case Western Reserve University, Cleveland, OH, United States

Keywords: cohesive gesture, co-speech gesture, reference resolution, preference, contrast, discourse, multimodal communication, moral issues

OPEN ACCESS

Approved by:
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

****Correspondence:***
Jennifer Hinnell
jennifer.hinnell@ubc.ca

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

Received: *04 February 2021*
Accepted: *05 February 2021*
Published: *05 March 2021*

Citation:
Hinnell J and Parrill F (2021) Corrigendum: Gesture Influences Resolution of Ambiguous Statements of Neutral and Moral Preferences. Front. Psychol. 12:664194. doi: 10.3389/fpsyg.2021.664194

A Corrigendum on

Gesture Influences Resolution of Ambiguous Statements of Neutral and Moral Preferences
by Hinnell, J., and Parrill, F. (2020). Front. Psychol. 11:587129. doi: 10.3389/fpsyg.2020.587129

In the original article, there was a mistake in **Figure 1** as published. **Figure 1** contains images reproduced from a television program that are part of the Red Hen dataset, available at redhenlab.org. As Frontiers does not apply Fair Use, the image has been removed from the figure, and a link has been inserted for readers to view the image on a public website. The corrected **Figure 1** appears below.

Additionally, in the original article, Uhrig (2020) was not cited. The citation has now been inserted in **Figure 1**, as shown above, and the corresponding reference has been added to the reference list.

The authors apologize for these errors and state that they do not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Uhrig, P. (2020). Multimodal Research in Linguistics. *Zeitschrift für Anglistik und Amerikanistik* 6, 345–349. doi: 10.1515/zaa-2020-2019

(1) *But apparently they like him on both sides. On the one hand, he believes in climate change so they consider him a Liberal. On the other hand, he doesn't believe in abortion so they consider him a Republican…*

FIGURE 1 | Contrastive use of gesture. 2015-09-24_1700_US_KABC_The_View, 191-201. Red Hen dataset http://redhenlab.org (click here or scan QR code to view the video clip; Uhrig, 2020).

HYPOTHESIS AND THEORY
published: 11 February 2021
doi: 10.3389/fpsyg.2020.552533

When Gesture "Takes Over": Speech-Embedded Nonverbal Depictions in Multimodal Interaction

Hui-Chieh Hsu, Geert Brône and Kurt Feyaerts*

Department of Linguistics, Faculty of Arts, University of Leuven, Leuven, Belgium

The framework of depicting put forward by Clark (2016) offers a schematic vantage point from which to examine iconic language use. Confronting the framework with empirical data, we consider some of its key theoretical notions. Crucially, by reconceptualizing the typology of depictions, we identify an overlooked domain in the literature: "speech-embedded nonverbal depictions," namely cases where meaning is communicated iconically, nonverbally, and without simultaneously co-occurring speech. In addition to contextualizing the phenomenon in relation to existing research, we demonstrate, with examples from American TV talk shows, how such depictions function in real-life language use, offering a brief sketch of their complexities and arguing also for their theoretical significance.

Keywords: depiction, multimodality, gesture, iconicity, embedding

OPEN ACCESS

Edited by:
Autumn Hostetter,
Kalamazoo College, United States

Reviewed by:
Gabrielle Hodge,
University College London,
United Kingdom
Jennifer Hinnell,
University of British Columbia,
Canada

****Correspondence:***
Hui-Chieh Hsu
huichieh.hsu@kuleuven.be

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

Received: *16 April 2020*
Accepted: *23 October 2020*
Published: *11 February 2021*

Citation:
Hsu H-C, Brône G and Feyaerts K (2021) When Gesture "Takes Over": Speech-Embedded Nonverbal Depictions in Multimodal Interaction. Front. Psychol. 11:552533. doi: 10.3389/fpsyg.2020.552533

INTRODUCTION

We communicate meaning to each other in different ways: by creating a physical analog, by relating ourselves to the physical world, or by assigning a sign to the meaning (Peirce, 1932; Clark, 1996; see also Goodman, 1968). The communication is in turn carried out in different channels: through speech, through nonverbal channels — such as manual gesture, eye gaze, and vocalization — or through a combination of multiple different channels. Linguists have long focused on how meaning is communicated through speech, primarily when it comes to the use of signs symbolically imposed on meanings, but also where the speaker makes meaning by anchoring themselves to the environment. With the multimodal turn in linguistics, as well as the resulting revitalization of interests in iconicity (Jakobson, 1966; Haiman, 1983; Simone, 1995; Wilcox, 2004; Perniss et al., 2010; Mittelberg, 2014), researchers have also examined how the speaker employs nonverbal signals alongside speech, such as co-speech iconic gestures (McNeill, 1992; Kendon, 2004; Cienki and Müller, 2008; Streeck, 2009) and pointing that accompanies verbal indices (Clark, 2003; Goodwin, 2003; Kita, 2003; Mondada, 2014; Langacker, 2016; Ruth-Hirrel and Wilcox, 2018). The recognition of signed languages as full-fledged linguistic systems likewise prompted curiosity about how nonverbal signals are used and coordinated to carry out the functions spoken languages serve (e.g., Stokoe, 1960; see also Vermeerbergen, 2006; Müller, 2018). While the current state of the art is a long way from the near-exclusive focus on symbolic signs at the onset of modern linguistics, the puzzle is not complete. Among the missing pieces are cases where meaning is communicated through iconic, nonverbal signals, in the absence of simultaneously co-occurring speech. Having only been explored in a handful of studies, this topic remains largely uncharted territory in the linguistics literature.

To contextualize, as well as better understand, phenomena that fall within this overlooked domain, we turned to the framework of language use proposed by Clark (1996) as a starting point. Building on Peirce's (1932) trichotomy of signs — icons, indices, and symbols — Clark distinguishes three methods in which meaning is signaled in language — depicting, indicating, and describing (as) — contextualizing the semiotic triangle in present-day linguistics. In a recent paper, Clark (2016) proposes the theoretical framework of the staging theory, in which he further elaborates on depicting as a basic method of communication. With examples from empirical data, he shows how depictions are employed in interaction to serve communicative functions, singles out numerous analytical dimensions that may prove crucial for the understanding of depictions, and, importantly, argues for the relevance of depicting to the study of language use that is on a par with indicating and describing.

Specifically, Clark (2016, pp. 324–325) defines depictions as iconic physical scenes people create and display, with a single set of actions at a single place and time, for the addressee to use in imagining the scenes depicted. They are physical analogs people produce, for the purpose of communicating meanings to which the analogs bear perceptual resemblance. Given the array of articulators the speaker is equipped with, depictions draw on various resources across different modalities, including manual gesture, bodily posture, head movement, facial expression, eye gaze, onomatopoeia, vocalization (Clark, 2019), any other "visible bodily action" (Kendon, 2004) — or even more broadly, any other "publicly intelligible action" (Mondada, 2019).

Depiction in Interaction

Found in an episode of *The Tonight Show Starring Jimmy Fallon*, the following example illustrates what a typical depiction in real-life language use looks like. In the excerpt, Kaley Cuoco, the talk show guest, recounts her experience of doing a "canyon swing."[1]

(1) Kaley Cuoco explains what canyon swing is: ". . . and you're supposed to just walk off, and it's a six-second free fall, and (*swings right arm back and forth, parallel to frontal plane*)[a] <u>then you swing</u>, for ten minutes."[2]

— *The Tonight Show Starring Jimmy Fallon*[3]

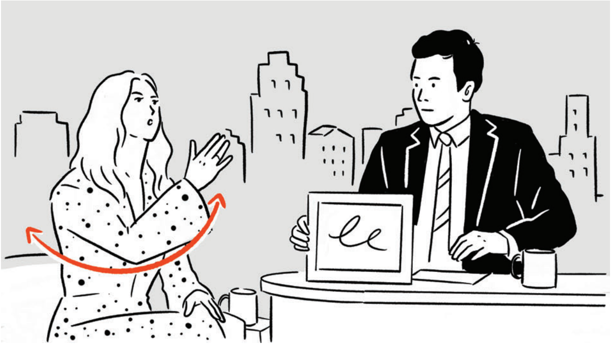


FIGURE 1 | Depiction in (1).

In addition to verbally describing the event of swinging by uttering *you swing*, Cuoco also stages a depiction simultaneously. Specifically, she deploys and coordinates a set of actions, consisting among others of a pendulum-like movement of her entire right arm. In a highly schematic fashion, the actions abstract from the depicted event of the actual canyon swing as she experienced it, to which the actions are iconic, with Cuoco's right arm being mapped onto the string of the canyon swing, and her right hand modeling her own body on the canyon swing.[4] The result is a depiction which, within the interpretive framework (see e.g., Bloom, 2010) that is the local context of language use in the exchange between Cuoco and Fallon, manifests physical resemblance to the depicted scene of the canyon swing.

By creating and displaying the depiction, Cuoco provides her audience — in this case Fallon, the audience in the recording studio, and the audience of the show as broadcast — with rich semiotic resources with which to imagine and comprehend the canyon swing scene, in a way that is concrete and perceptually tangible: Normally, with the descriptive verbal phrase *you swing* alone, the audience imagines the swinging based primarily on the symbolic (and therefore largely arbitrary) form-meaning relation associated with the verbal phrase. With the aid of the depiction, the audience is afforded additional semiotic resources with which to imagine and therefore understand the swinging — including the manner, direction, and speed — in a more direct, albeit highly schematic, manner, as the link between the form of the depiction and its meaning is iconically motivated.

A Schematic Vantage Point

Essentially, depictions, as defined by Clark within the staging theory, make up cases of language use where the relation between the semiotic signal and its denotation is iconic, contrasting with

[1] In line with Clark's (2016) notation, nonverbal signals are described in italics. Nonverbal signals in brackets do not co-occur with speech; those that do co-occur with speech are in parentheses, their co-occurring speech underlined. Where drawings are included, the superscript letters in the token description indicate the corresponding drawings.

[2] Not all actions observed in the tokens are addressed in the present paper, for theoretical reasons (see section "Speech-Embedded Nonverbal Depictions") as well as methodological ones: Cuoco, for instance, can be observed displaying some facial expression at the same time the depiction in (1) is being staged. While the facial expression may be depictive of her facial expression during (or, more likely, around) the time of the depicted event, it is equally plausible that it is a display of her stance toward the fact that she was tricked by her fiancé into doing something as scary as a canyon swing, a piece of information she has revealed in prior discourse. In this sense, the facial expression is, at best, peripheral to the depiction of the canyon swing, if part of any depiction at all. Significant in their own right (see e.g., Tabacaru, 2014; Janzen, 2017), such actions fall beyond the focus of the present paper that is prototypical depictions, therefore not included in the token description.

[3] The drawings included in the present paper are made by Yuga Huang (yugagagahuang.myportfolio.com), based on screenshots taken from four TV talk shows (see section "Methods"): *The Ellen DeGeneres Show*, *Late Night with Seth Meyers*, *Conan*, and *The Tonight Show Starring Jimmy Fallon*.

[4] As is the case for any depiction, the iconic relation between Cuoco's actions and the canyon swing scene is a complicated one, involving not only physical resemblance between the articulators and the depicted scene, but also contextual information, the embodied encyclopedic knowledge of Cuoco and her audience, and, importantly, an array of complex metonymic mappings (see Arnheim, 1969; Mittelberg and Waugh, 2014). Given the focus of the present paper, the detailed mechanisms of metonymy in the examples are discussed concisely and selectively.

descriptions — whose form-function relation is symbolic — and indications — whose form-function relation is indexical (Peirce, 1932; Clark, 1996, 2016). Importantly, the three ways in which meaning can be signaled are not mutually exclusive. It is possible, and indeed often the case, for a communicative form to signal meaning in more than one way: A depiction of someone finger-pointing at something, for example, is both depictive and indicative; conventionalized ideophones such as *meow* and *whack* are descriptive as well as depictive (Dingemanse, 2017a); likewise, depicting constructions in signed languages exhibit properties associated with both descriptions and depictions (Ferrara and Hodge, 2018). Depicting, indicating, and describing are therefore better considered, not as discrete categorical notions, but as properties or dimensions of communicative signals, a view that finds advocates in more recent studies (McNeill, 2005; Mittelberg and Evola, 2014). In this sense, a prototypical depiction is really a communicative signal whose depictive property is more salient than its indicative and descriptive properties.

While Clark's approach to depictions may be new to the field, the notion of depicting itself is not, and neither are the plethora of phenomena that fall within Clark's definition of depicting (though many of these have been marginalized in the literature, see subsection "An imbalance in the Literature" and Dingemanse, 2017b). The very term of depicting has been used by a great number of researchers — some in more clearly delimited senses than others — to refer to various different subsets of iconic, nonverbal strategies of communication. Examples of a more systematic use of the term depicting can be found in the research of Müller and Streeck. Drawing on an analogy to the techniques employed by visual artists, Müller (1998b) identifies four basic techniques of gestural depiction: acting, molding, drawing, and representing, later breaking them down into two: acting and representing (Müller, 2014). Observing how a single object can be depicted by the hand in multiple different ways, she explores the interplay between gestural representation and conceptualization. Streeck (2009), on the other hand, views gestures as organic products of humans acting in the material world as well as in interaction with each other. Examining empirical data in a "micro-ethnographic" fashion, he identifies, heuristically, a dozen depiction methods, positing that "to depict a phenomenon means to analyze and represent it in the terms that the given medium, communicative modality, or symbol system provides" (Streeck, 2008, p. 286).

Also covered by Clark's notion of depicting are manual gestures with an iconic form-meaning relation, a topic that has been explored by a great number of researchers, though not necessarily using the term depicting (e.g., Calbris, 1990; McNeill, 1992, 2005; Gullberg, 1998; Kendon, 2004; Sowa, 2006; Cienki and Müller, 2008). Depending whether the denotation is something concrete in the material world, these gestures are often divided into two separate groups: "iconic" (or "representational," "referential," "imagistic") and "metaphoric" (or "conceptual," "ceiving") (see the reviews in Kendon, 2004; Streeck, 2008; Mittelberg and Evola, 2014). On a more schematic level, Clark's notion of depicting covers also phenomena which are often approached separately and independently, but which share the same defining property of iconicity. These include topics such as quotation, demonstration, enactment, pantomime, mimesis, facial gesture, ideophone, constructed action, and depicting construction (e.g., Mandel, 1977; Clark and Gerrig, 1990; Chovil, 1991; McNeill, 1992; Wade and Clark, 1993; Kita, 1997; Gullberg, 1998; Liddell, 2003; Kendon, 2004; Zlatev, 2005; Taylor, 2007; Cienki and Müller, 2008; Vandelanotte, 2009; Cormier et al., 2012; Dingemanse, 2013; Cormier et al., 2016; Gärdenfors, 2017). Furthermore, depictive semiotic signals are ubiquitous not just across said topics, but are prevalent across modalities, and observed across communicative ecologies (such as hearing to hearing, and deaf to deaf; see Ferrara and Hodge, 2018).

As the above overview shows, phenomena in language use that pertain to iconicity have been explored by many, from various perspectives and in different theoretical frameworks. For the present study, Clark's (2016) account of depicting was chosen as the starting point through which to explore the aforementioned oversight in the literature, for the reason that Clark's definition of depicting is a well delimited one, but more importantly, because it provides a schematic vantage point from which to approach iconicity. Rather than a mere change of terminology, the framework situates existing research in a bigger picture, uniting and consolidating numerous research traditions. This affords the researcher the possibility of observing iconic language use on a more schematic level, and, in turn, the potential to identify patterns that have hitherto eluded scholarly attention. Indeed, some early findings using Clark's framework of depicting have already been reported, for both spoken and signed language interactions (e.g., Ferrara and Hodge, 2018; Hodge et al., 2019; Hsu et al., to appear). Given that this framework was put forward only fairly recently, it is reasonable to expect more fine-grained analyses in the near future.

The Present Study

In the following sections, we start from one of the central theoretical distinctions in Clark's framework, namely the four-way typology of depictions (section "Clark's Typology of Depictions"). A closer examination reveals potential insufficiencies of the typology, leading to problems in categorization. In light of this, we tap into a corpus of American TV talk shows that we constructed specifically for this purpose (section "Methods"). Through systematic data annotation and operationalization of relevant theoretical notions in Clark's framework, we establish a methodology for researching depictions that is both empirically grounded and theoretically valid. Confronting Clark's proposed typology with real-life usage events taken from the corpus (section "Depiction Type Attribution"), we zero in on the issues encountered in depiction type attribution, pinpointing underspecification and form-function conflation as the major underlying causes. This critical examination further leads to an alternative, gradient conceptualization of Clark's original typology. Crucially, this alternative conceptualization brings the aforementioned overlooked domain to the fore, namely cases where meaning is communicated iconically, nonverbally, and without simultaneously co-occurring speech. A review of existing studies then reveals a curious imbalance in the literature, between gesture employed with and without co-occurring speech: Ubiquitous in language use, cases of gesture without temporally overlapping speech have been widely acknowledged by researchers, but unlike those with cotemporal speech, they have not received proportionate scholarly attention.

In view of this, we zoom in on a subset of phenomena within this marginalized domain: "speech-embedded nonverbal depictions" — which we define in more technical terms as "depictions that are embedded in speech, but that are not depictions of non-depictive speech" (section "Speech-Embedded Nonverbal Depictions"). This definition excludes depictions of descriptive and indicative speech (e.g., canonical quotations), but takes into account cases of depictive speech (e.g., ideophones), allowing us to focus on depictions that have until recently been largely overlooked. Taking embedding in a strictly formal sense — in terms of temporal overlap — this approach also steers clears of the problems of Clark's original typology. Following a delimitation of speech-embedded nonverbal depictions, a brief sketch is offered of how such depictions function in naturally occurring discourse. With examples from the TV talk show corpus, we demonstrate the theoretical significance of speech-embedded nonverbal depictions in relation to current topics in the literature of relevant fields of inquiry, calling for further research on this marginalized topic along various dimensions.

Essentially, the aim of the present paper is first and foremost to establish a case for speech-embedded nonverbal depictions, by demonstrating their theoretical significance, but also by laying out the methodological groundwork for systematic empirical investigations. The findings — methodological, theoretical, and empirical — of the present study are therefore reported throughout sections "Methods," "Depiction Type Attribution," and "Speech-Embedded Nonverbal Depictions," although the more technical discussions of empirical tokens are concentrated in section "Speech-Embedded Nonverbal Depictions."

As the term "speech-embedded nonverbal depiction" suggests, the main argument of the present study builds in part on distinctions such as "verbal vs. nonverbal." The use of such dichotomous terms calls for clarification. On the technical level, modality and signaling method need to be teased apart. Like other modalities, speech can be depictive, indicative, descriptive, or any combination thereof. Since the focus here is on depicting, unless otherwise specified, we use "speech," as well as related terms and modifiers such as "verbal" and "utterance," as a shorthand term for non-depictive speech, that is descriptive or indicative speech, where speech is understood in a modality-agnostic (Dingemanse, 2019) sense compatible with both spoken and signed languages. The distinction is therefore really between different combinations of signaling method and modality (e.g., "non-depictive speech vs. depictive signals"), and not between different modalities. The specific use of the terms serves the purpose of naming, and therefore tackling, the specific phenomena in question, rather than asserting rigid categorical boundaries based on a dichotomy between "language proper" and "paralinguistic noise" such as gesture. In line with most researchers in relevant fields of inquiry, we view all communicative behavior that is deemed (intentionally) meaningful (see Kendon, 2004) as integral to talk, to speaking, and to language use. It follows that the seemingly discrete categorical notions are really heuristics that guide the recognition of phenomena along the messy and overlapping continua that constitute language use. As Streeck (2009, p. 11) also acknowledges, categorization "helps us organize our analysis, [. . .] reminds us of the wide range of different uses to which gesture is put, and thus keeps us from drawing overly broad generalizations from a narrow data-set."

CLARK'S TYPOLOGY OF DEPICTIONS

Based on the functional relations between depictions and their adjacent or accompanying (non-depictive) speech, Clark (2016) puts forward a typology consisting of four types of depictions: adjunct (where the depiction, acting like a nonrestrictive modifier, co-occurs with and illustrates descriptive speech), indexed (where the depiction is indexed by an indexical device in speech), embedded (where the depiction takes up a syntactic slot in a descriptive verbal utterance), and independent (where the depiction stands alone). Cuoco's depiction in (1), repeated in (2), is an example of an adjunct depiction, as her depiction co-occurs with, and illustrates, the descriptive verbal phrase *then you swing*, thereby adding to it iconic imagistic details of the event, such as the physical configuration of the swing, and the manner and directionality of the movement, though only schematically. The depictions in (3)–(5), which are manipulated variations based on (2), illustrate the other three types of depictions in Clark's typology.

(2) Adjunct depiction: ". . . and you're supposed to just walk off, and it's a six-second free fall, and (*swings right arm back and forth, parallel to frontal plane*) <u>then you swing</u>, for ten minutes."

(3) Indexed depiction: ". . . and you're supposed to just walk off, and it's a six-second free fall, and then you swing like (*swings right arm back and forth, parallel to frontal plane*) <u>this</u>, for ten minutes."

(4) Embedded depiction: ". . . and you're supposed to just walk off, and it's a six-second free fall, and then you [*swings right arm back and forth, parallel to frontal plane*], for ten minutes."

(5) Independent depiction: Jimmy Fallon: "How does a canyon swing work?" Kaley Cuoco: "[*swings right arm back and forth, parallel to frontal plane*]"

As the depiction in (3) is connected to the rest of the utterance by the verbal demonstrative *this*, it is categorized as an indexed depiction (but see subsection "Embedding"). Importantly, indexed depictions differ from most deictic expressions, in that the referent of the indexical device is not something that already exists (either physically or conceptually), but the depiction created by the speaker for local purposes. In (4), the depiction is embedded in speech, in the sense that it fills the syntactic slot where a verbal phrase (e.g., *swing in the canyon*) otherwise would; it is therefore an embedded depiction. In (5), Cuoco answers Fallon's question not with descriptive speech, but with a set of nonverbal, depictive actions, which stands alone and contributes to the discourse independently, hence an independent depiction.

The categorization might, at first sight, appear to nicely capture the various possible functional relations between depictions and speech; however, it really only is the case with tokens that are prototypical exemplifiers of the four depiction types. Upon careful consideration, ambivalence surfaces, in gray areas where the categories overlap. For instance, if the depiction in (2) did not co-occur with, but followed the phrase it illustrates (*then you swing*), would it still count as an adjunct depiction, and not as an independent one? Is an embedded depiction that takes up a syntactic slot on the clausal (or even sentential) level still to be categorized as an embedded depiction, and not as an independent one? Such issues only become exacerbated once the typology is confronted with the messy, heterogeneous tokens in empirical data.

While Clark's typology is likely put forward, not as a definitive assertion about discrete categories, but in a heuristic way in which to demonstrate the diverse depiction-speech relations, it was taken as the starting point for our empirical investigation on depicting. In what follows, we critically scrutinize the four depiction types as defined by Clark in a bottom-up fashion, comparing them to the empirically attested phenomena observed in a corpus constructed for this purpose — a process through which previously overlooked issues can be identified and addressed, potentially leading to an understanding of depicting that is more well-rounded, both theoretically and empirically.

METHODS

The data examined for the present study comprise video recordings of American TV talk shows, which were chosen for a number of reasons: While the topics of the talk show episodes may be predetermined, there is nonetheless a high level of spontaneity in the way the topics are actually delivered or discussed by the hosts and guests. Video recordings of talk shows are plentiful, easy to collect, and come in good quality. In addition, the unbalanced interpersonal power dynamics found in certain settings (e.g., in contexts of instruction, see Goldin-Meadow, 2003; Hsu et al., to appear) are to a large extent absent. Indeed, a growing number of studies on multimodal communication have examined television data for similar reasons [see the studies drawing on the databases of the Red Hen Lab (www.redhenlab.org) and the TV News Archive (archive.org/details/tv); e.g., Steen and Turner, 2013; Winter et al., 2013; Zima, 2017; Hinnell, 2018, 2019]. On top of all these, TV talk shows abound in recounts of past experiences and enactments of hypothetical scenarios, both of which contribute to the richness of depicting.

Corpus and Annotation

To avoid generalization over individual idiosyncrasies, a corpus of video recordings was constructed, comprising video clips randomly retrieved from the official YouTube channels of four American TV talk shows: *The Ellen DeGeneres Show*, *Late Night with Seth Meyers*, *Conan*, and *The Tonight Show Starring Jimmy Fallon*. Specifically, we examined only segments of host-guest interaction — that is, where the host and guest(s) are both physically present in the recording studio, and where they interact with one another — as these segments approximate canonical, spontaneous face-to-face interaction more so than other types of "interaction" on TV (e.g., where the host speaks directly into the camera, see Turner, 2017). In total, 147 video clips were examined, amounting to a total duration of approximately 10 h 37 min.

The video data were imported into ELAN,[5] where tokens of depictions were identified and segmented (see subsection "Unit of Analysis" for segmentation). Aware of the problems of form-function conflation such as circularity (see e.g., Croft, 2001), a strict line was drawn between the formal and functional properties of depictions. Given the complexity and heterogeneity of depictions, for each of the tokens, we describe the salient form features (features of "articulator form," rather than "gesture form"; see Hassemer, 2016) of the actions that are core to the depiction (the "modality-agnostic stroke"; see subsection "Unit of Analysis"). Specifically, McNeill's (1992) gesture space is referenced in the description of location. For other parameters such as articulator shape, movement, and orientation, the annotation is informed by Bressem's (2013) form-based annotation scheme for manual gestures, which also steers clear of any functional interpretations. On top of the description of depictions per se, our annotation also includes, among others, depiction type (in Clark's typology), immediately adjacent or overlapping speech, grammatical level of embedding, as well as parameters pertaining to depiction-speech relations. A screenshot of our full annotation in ELAN (which includes annotation tiers beyond the scope of the present paper) can be found in the **Appendix**. In total, 217 tokens of our target phenomenon — speech-embedded nonverbal depictions — were identified and annotated in our corpus (see section "Speech-Embedded Nonverbal Depictions" for a full definition of such depictions), providing the empirical basis of the present study.

Due to the limitations of format — in the sense that it is not possible to include dozens of video clips with sufficient length to cover all relevant contextual information — the descriptions of the tokens in the present paper were adapted accordingly, to facilitate the reader's understanding and imagination of the actions described. In addition to overly trivial details being omitted, functional descriptions are supplemented where a purely form-based description would be overly lengthy and confusing. These functional descriptions are always marked and preceded by the phrase *as if*.

Unit of Analysis

The construction of our corpus, specifically our attempt at systematic annotation, was not without challenges. Among them is the lack of a readily operationalizable unit of depiction. In his account of depicting, Clark (2016) does not spell out the segmentation of depiction tokens (neither does Müller, 2014 or Streeck, 2008), which is essential to the establishment of the basic unit of analysis, and therefore to systematic annotation. While clear-cut examples such as (1) do exist, more often than not, a depiction is preceded or followed, with or without

[5] ELAN (Version 5.2) [Computer software]. (2018, April 04). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

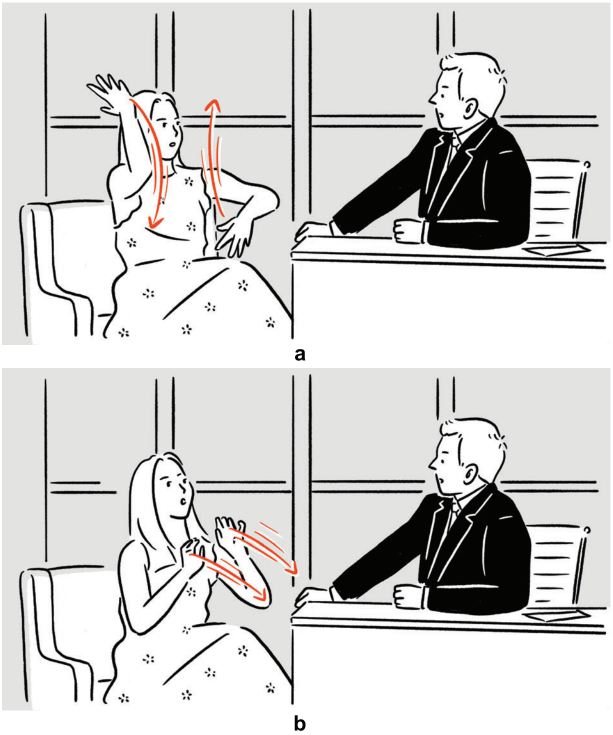


FIGURE 2 | Depictions in (6).

speech "intervening," by another depiction, which can be either similar or distinct in form and meaning, as illustrated in the following examples. For clarity, they are presented with our final segmentation, explained immediately below.

(6) Lauren Ambrose on backstage costume change on Broadway: "I mean sometimes it's like twenty seconds, for like, full-on, [*vocalizes whistle-like* fsss *sound, moves both hands vertically, fingers spread, in opposite directions, in front of head and torso*][a] — [*vocalizes whistle-like* ffft *sound, gazes at the front, into the distance, moves both hands along sagittal axis away from body, fingers spread, palms away from body*][b]."

— *Late Night with Seth Meyers*

(7) Tracy Morgan explains what bingo wing is: "When an old woman hits bingo, she goes, [*vocalizes* bingo,[6] *raises and shakes both arms, elbows bent*][a], and then [*raises and shakes left arm, left elbow bent; moves right hand, fingers spread, back and forth under and perpendicular to left arm*][b], bingo wings, [*raises and shakes both arms, elbows bent*][c] (.) [*raises and shakes both arms, elbows bent*][d]."[7]

— *Conan*

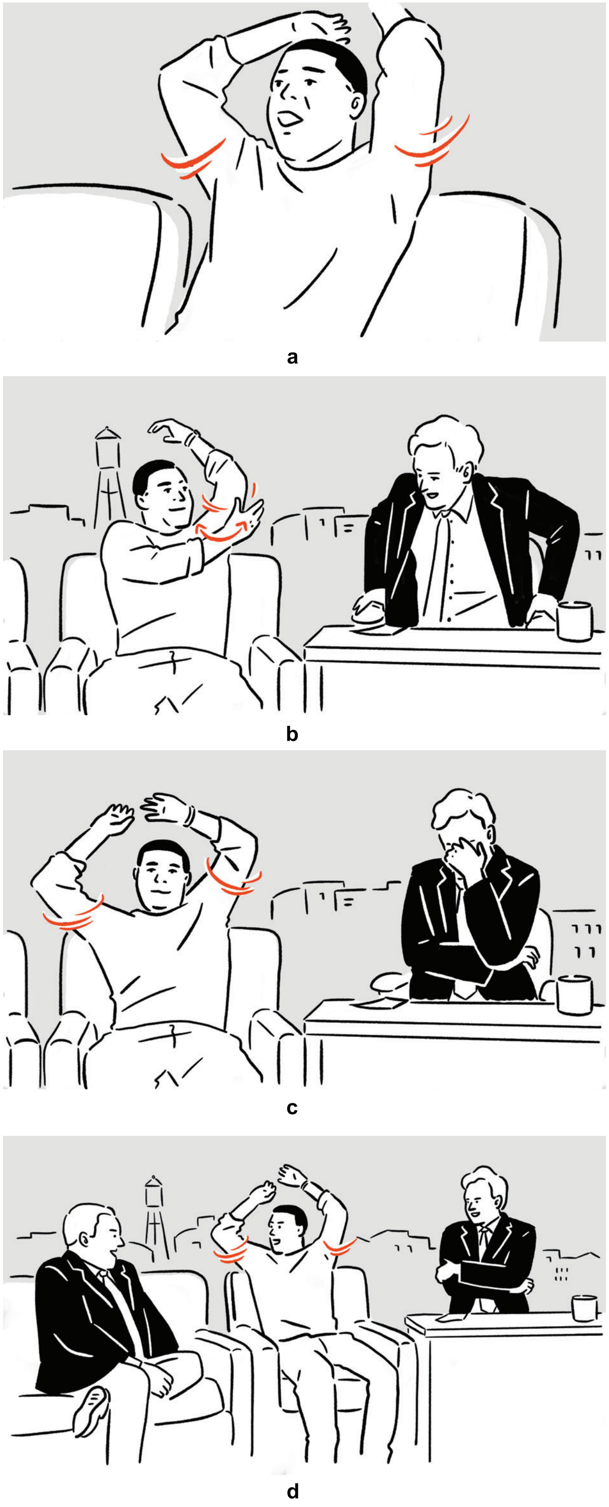


FIGURE 3 | Depictions in (7).

In (6), Ambrose recounts her experience of performing on Broadway, specifically the backstage costume change operations which she finds unbelievably fast. In the first set of actions,

[6] The word *bingo* is used in this specific case as a conventionalized ideophone, therefore categorized not as descriptive speech, but as a depictive signal (see section "Speech-Embedded Nonverbal Depictions").

[7] Following the practice in Conversation Analysis, we use (.) to indicate a short pause or micro-pause, shorter than 500 ms (see Mondada, 2016).

with her moving hands standing for the hands of the multiple members of backstage personnel quickly working on her clothes and makeup, and with the rest of her body portraying herself in the depicted scene, the highly efficient change of costumes is depicted[8]; in the second set, she depicts someone already pushing her, with their hands, back to the stage. The two sets of actions are deployed consecutively, without a pause. They utilize the same channels of communication (mainly, vocalization and manual movements), but the actions are distinct in form (*fsss* vs. *ffft*, vertical movement vs. movement along the sagittal axis). At the same time, the two sets of actions are functionally interrelated, as they each depict a part of a larger sequence of events. In (7), Morgan explains the (folk) etymology and concept of "bingo wings" — the flabby triceps area that wobbles as the arm moves — through four sets of actions, which involve him shaking his own arm in an exaggerated manner so as to make the triceps shake, whilst vocalizing *bingo*. Here, the four sets of actions are "interrupted" by speech and a pause, but they are very similar in both form and meaning. In fact, all except for the second set are essentially iterations of the same actions. These two tokens exemplify the commonly observed mismatch in terms of sequentiality, form, and meaning — sequentially consecutive depictions, for instance, can be distinct in form but interrelated in meaning, while depictions that are separated by descriptive words can be identical to each other both in form and meaning — posing challenges to systematic segmentation.

While there is probably no universally valid definition of a unit of depiction, to ensure consistency in annotation, we adapted and operationalized the notion of the gesture phrase as the basic unit of depiction for the present study. The gesture phrase, as defined by Kendon (1972, 1980) primarily for the study of manual gesture, consists of the preparation phase, the stroke, and any subsequent sustained position. Given the fact that depictions often make use of modalities other than manual gesture, we schematized from Kendon's definition, making the gesture phrase a modality-general — or, following Dingemanse (2019), "modality-agnostic" — notion, where the stroke can be carried out by any possible articulator, or combination of articulators. In this sense, a unit of depiction consists of a stroke of action (be it manual gesture, vocalization, head tilt, leg or torso movement, etc.) as its core, with its start marked by the onset of the preparation phase of the action, its end either by a complete rest, or by the onset of another modality-agnostic gesture phrase. The operationalization of the gesture phrase as a modality-agnostic notion is in line with recent works on comparable topics (e.g., Ferrara and Hodge, 2018; Dingemanse, 2019), but also motivated by Kendon's (2004) view of gesture as "visible bodily action," as well as Mondada's (2019) notion of "publicly intelligible action." Adopting a broader sense of the term "gesture" that is not limited to manual actions, we take into consideration all nonverbal signals that contribute to the meaningfulness of the depictions in question, including visual but also auditory ones.

It is in this way that the depictions in (6) and (7) were segmented, as indicated by the brackets above. Despite the two depictions in (6) being staged back to back, and despite their shared semantic thread, they exhibit two distinct sets of actions, with two distinct strokes of actions (both of which with simultaneous utilization of vocalization and manual gesture), rendering them not one but two units of depiction. In (7), the four sets of actions share many common features, with the fourth being a reiteration of the third. However, since each of them is followed by either a complete rest or another gesture phrase, they make up four gesture phrases, and therefore four units of depiction in our annotation. All other tokens in our corpus were segmented following the same principle.

DEPICTION TYPE ATTRIBUTION

Confronted with our TV talk show data, Clark's staging theory indeed captures much of the complexities of depictions rather intuitively and coherently, especially in the identification of depictive properties in communicative signals. At the same time, however, this process also revealed potential insufficiencies. In addition to methodological issues such as segmentation, also foregrounded are problems on a more theoretical level, including the aforementioned issue of depiction type attribution.

As mentioned in section "Clark's Typology of Depictions," Clark's (2016) definition of depiction types leaves gray areas for non-prototypical cases. This is confirmed by our attempt at imposing the typology on our corpus data. Some of the frequently encountered challenges are illustrated by the following example.

(8) Tracy Morgan on the quality of his facial muscles: "Yeah, I'm your rubber-band man, [*vocalizes* brbrbrbr *sound,*[9] *shakes head sideways quickly, causing facial muscles to vibrate accordingly*][a]."[10]

— *Conan*

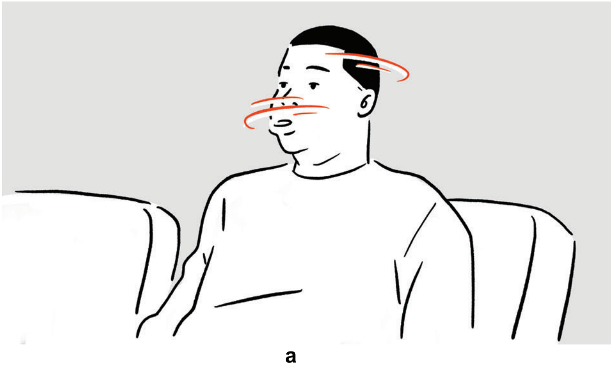


FIGURE 4 | Depiction in (8).

[8] With the speaker taking up the roles of multiple actors, these actions also instantiate what Clark (2016) identifies as an actor-actor hybrid depiction.

[9] The *brbrbrbr* sound is an ideophone, albeit a non-conventional one. Depictive rather than descriptive (see section "Speech-Embedded Nonverbal Depictions"), it is a part of the depiction.

[10] If the reader finds the Tracy Morgan quote confusing, that is because it is meant to be confusing, as evidenced by what the host, Conan O'Brien, remarks immediately after, "it's the most sense you've made all night."

Following the verbal phrase *rubber-band man*, Morgan depicts the elastic, rubber-like quality of his skin, by shaking his head violently so that the cheeks wobble, thereby illustrating, metonymically, the verbal phrase.

As defined by Clark (2016, p. 326), adjunct depictions are the ones that are "timed to overlap with" their verbal affiliates to which they are adjoined, so as to elaborate on them "as if they were non-restrictive relative clauses" or nonrestrictive modifiers. In (8), as the brackets indicate, there is no temporal overlap between the speech and the depiction, although at the same time, the depiction elaborates on its verbal affiliate *rubber-band man*, albeit metonymically, rendering it unclear whether it is an adjunct depiction. If, for the sake of discussion, we do not categorize it as an adjunct depiction, for the reason that it does not share all the properties of a prototypical adjunct depiction, further issues arise: Does it belong to embedded depictions, which "function as parts of utterances — as if they were words, phrases, or other segments," or to independent depictions, which make "independent contributions to the discourse" (Clark, 2016, pp. 325–326)? In other words, is the depiction embedded in the verbal sentence as if it were an apposition, therefore a part of the utterance, or does it make a contribution that is independent of the preceding utterance (but see subsection "Embedding" for the issue of independence; see also Lehmann, 1988; Hodge and Johnston, 2014 on comparable phenomena observed in other communicative ecologies)? On top of that, how "independent" is "independent" enough? What kind of independence is at issue: syntactic, semantic, or something else? These questions suggest calibration may be needed before the typology can be applied empirically.

Indeed, a critical review of the typology brings to light two major causes of confusion: underspecification and form-function conflation. Underspecification is most manifest where independent and embedded depictions are concerned — it is unclear what level, and what kind, of independence is sufficient for a depiction to be categorized as "independent." Similarly, for indexed depictions, it is not specified whether they include only depictions indexed by indexical pronouns (e.g., *this* in *I'd do it like this*), or also those indexed by indexical modifiers (e.g., *that* in *they chose that color*), despite the fact that indexical pronouns and indexical modifiers are indexical in distinct ways (see subsection "Embedding").

Form-function conflation, on the other hand, is a problem that is inherent in the typology itself. Although the four types of depictions are, as Clark puts explicitly, defined in terms of their discourse functions, both formal and functional criteria are present in their definition. Take the aforementioned case of adjunct depictions. While it is indeed a functional definition that an adjunct depiction elaborates on its verbal affiliate in a way that is similar to a non-restrictive relative clause, the criterion that an adjunct depiction is "timed to overlap with" (Clark, 2016, p. 326) its affiliate is unequivocally a formal one. Conflation is also found among different functional notions. For example, as indexation and embedding are not two mutually exclusive functional concepts, ambiguity often surfaces where an indexed depiction is itself part of an embedded depiction. In fact, mutual inclusion can, strictly speaking, be found among all of the canonical functions associated with each of the depiction types — elaboration, indexation, embedding, and independent meaning contribution. Issues such as these call for thorough reconsideration of depiction categorization in relation to speech.

Typology of Depictions Reconceptualized

Serving as the starting point for the present study, the critical examination of Clark's (2016) typology presents, more importantly, an analytical process toward a better understanding of depictions. Among the results of this process is a reconceptualization of the depiction types, which in fact foregrounds some of the implicit insights of Clark's original typology. In this subsection and the next, we consider the theoretical implications of this reconceptualization, visualized in **Figure 5**, before tackling the issues of the typology raised above.

The four depiction types are placed along a continuum, with varying levels of information contribution from two different combinations of modality and signaling method: non-depictive speech (i.e., indicative and descriptive speech) and depictive signals (e.g., depictive manual gesture, depictive bodily movement, depictive speech), where speech is understood in

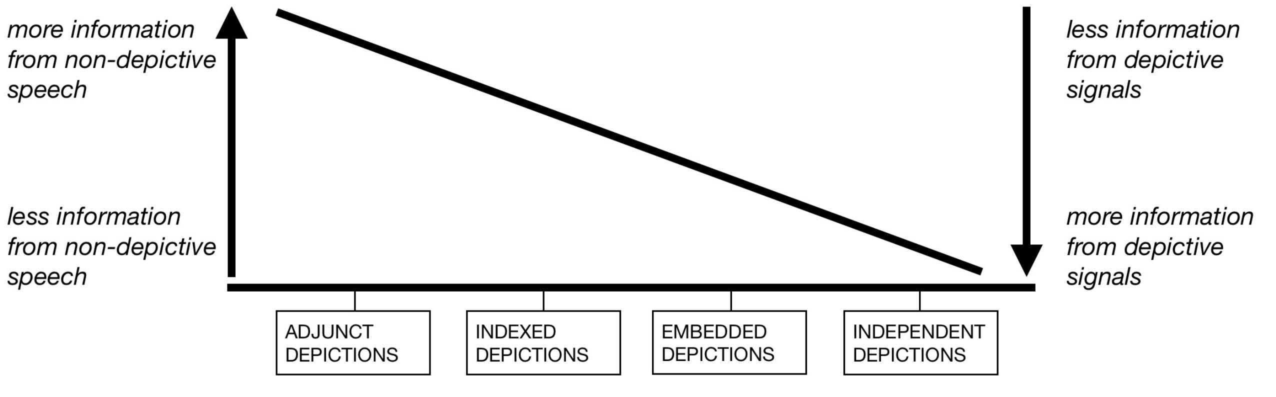

more information from non-depictive speech
less information from non-depictive speech
less information from depictive signals
more information from depictive signals
ADJUNCT DEPICTIONS
INDEXED DEPICTIONS
EMBEDDED DEPICTIONS
INDEPENDENT DEPICTIONS

FIGURE 5 | Continuum of information contribution from non-depictive speech and depictive signals.

the above-mentioned modality-agnostic sense.[11] On the left half of the continuum are cases where more information is communicated through non-depictive speech, and where relatively less information comes from depictive signals. Here we find adjunct and indexed depictions: As adjunct depictions illustrate what is said in the descriptive speech they co-occur with, part of the composite meaning is conveyed through their co-occurring speech [cf. "composite utterance" (Enfield, 2009) and "multimodal attribution" (Fricke, 2008, cited in Bressem, 2014; Ladewig, 2020)]. In the case of indexed depictions, indicative speech provides essential deictic information, directing the addressee's attention toward the depiction, through which meaning is conveyed. With depictive signals communicating meaning that is relatively complementary to the non-depictive speech they accompany, this half of the continuum largely coincides with the scope of the research program on co-speech gesture.

For cases closer to the right half of the continuum, relatively more information is communicated through depictive signals, and relatively less information comes from non-depictive speech. Embedded and independent depictions are located on this side of the continuum: Embedded depictions (more precisely, the stroke phase thereof, see subsection "Embedding") convey meaning without the accompaniment of temporally co-occurring non-depictive speech, but are formally and functionally framed by the non-depictive speech surrounding the syntactic slot that they fill. Independent depictions also convey meaning without simultaneous non-depictive speech, and do so, according to Clark's (2016) definition, independently of preceding or following speech. Without temporally overlapping non-depictive speech, depictive signals on this half of the continuum often contribute to the discourse essential information that is absent in the adjacent speech. In the sense that these are cases where depictive signals fill in temporal slots in the discourse, they are, in more general terms, cases of iconic gesture without co-occurring speech.

Thus conceptualized, the four depiction types as defined by Clark, of which the prototypical cases can be located as four points along the continuum, really capture the different levels of "division of labor" between non-depictive speech and depictive signals — or more generally speaking, between speech and depictions. In some cases, speech takes up more of the "load" of meaning communication; in others, the depiction "takes over," showing meaning in iconically motivated ways. Importantly, the reconceptualization is not meant as a solution to the issues of the original typology. As the choice of term "continuum" suggests, it presupposes gradience rather than categoriality. Given the challenge of quantifying the amount of information communicated through speech as compared to depictions, the continuum is not one with strictly defined criteria either. Rather, it serves as a heuristic for identifying the varying levels of "trade-off" in terms of meaning contribution between non-depictive speech and depictive signals — not as dichotomous oppositions, but as two of the many sets of communicative resources available to the speaker in language use. It shows how, in staging different types of depictions, the speaker "packages" information in different ways, "distributing" it over speech and depictions, be they co-expressive, with or without "redundancy."

An Imbalance in the Literature

In addition to providing an alternative vantage point from which to consider the speech-depiction relations in the four depiction types identified by Clark (2016), the reconceptualization of the typology bears further theoretical relevance. Among other things, it brings to the fore an imbalance in the literature between studies on iconic gestures with and without co-occurring speech.

Largely coinciding with the left half of the continuum, where speech plays a relatively dominant role, and where adjunct and indexed depictions are located, the topic of iconic co-speech gesture has been core to modern gesture studies, with an extended body of dedicated research. Some scholars, for instance, have explored how gestures complement or supplement the semantics of their co-occurring speech (see the pioneering research by McNeill, 1992; Kendon, 2004); others have investigated how gesture and co-occurring speech package meaning in different ways ("imagistic" versus "linguistic"), debating how the two processes relate to each other (e.g., de Ruiter, 2000; Kita, 2000; McNeill and Duncan, 2000); still others have investigated the "deeper" link between gesture and co-occurring speech, as well as its implications in psychology and evolution (e.g., Stokoe, 2001; Arbib, 2005; Tomasello, 2008; McNeill, 2013a; Kita et al., 2017). Despite the late revival of the topic, formidable groundwork has been laid for the understanding of the workings of iconic co-speech gesture.

In contrast, phenomena that fall closer to the other end of the continuum — cases where iconic gesture communicates meaning without co-occurring speech, such as embedded and independent depictions — have not received equal attention. McNeill (2005, p. 5), for instance, identifies gestures that "occupy a grammatical slot in a sentence" as "speech-framed" or "speech-linked" gestures on Kendon's Continuum (see also Kendon, 1988a; McNeill, 1992), but does not include them in further discussion. This is echoed by the general trend in gesture studies. Iconic representational gestures, for instance, have been explored by many, but with most of the studies focusing primarily on those co-occurring with speech (e.g., Müller, 1998a; McNeill, 1992; Kendon, 2004; Cienki and Müller, 2008; Enfield, 2009; Streeck, 2009). Fricke (2012, 2013) in her research delves into what she calls multimodal attribution, where gestures provide supplementary and sometimes essential information, but with the presence of co-occurring speech (see also Bressem, 2014). Similarly, Mittelberg and Evola (2014, p. 1734) observe that "iconic gestures can be produced to fill a semantic gap in speech, especially when representing spatial imagery like size, shape, motion, or other schematic, partial images which take advantage of the affordances of gestures versus speech," but keep their focus on gesture-speech co-occurrence. Indeed, as Fricke points

[11] Incidentally, **Figure 5** bears resemblance to figures presented in Dingemanse (2017a) and Ferrara and Halvorsen (2017), which capture the continuum along which the dual semiotic properties — of being both depictive and descriptive — of ideophones and iconic lexical signs can be exploited. The resemblance is however only on the level of visualization: What **Figure 5** presents is not the interplay between the different semiotic properties within individual signals, but how depictive signals (signals whose semiotic properties are predominantly depictive) relate to non-depictive speech in the four types of depictions identified by Clark (2016).

out, research in gesture studies has not yet moved beyond "the *assumption* that [...] gestures can fill syntactic gaps in linear verbal constituent structures" (Fricke, 2013, p. 748; emphasis ours). As the continuum in **Figure 5** shows, however, to focus only on gestures with co-occurring speech is to miss out on the other half of the picture.

To date, only a relatively small number of researchers have tapped into iconic gesture without co-occurring speech in naturalistic language use (but see reports from experimental settings, e.g., Sambre et al., 2019). Fricke (2012, cited in Müller et al., 2013, p. 65), for instance, identifies two types of gesture-speech integration, arguing that "gestures may be integrated by positioning, that is either through occupying a syntactic gap or through temporal overlap; or they might be integrated cataphorically, that is by using deictic expressions." Though proposed for gestures in general, this distinction shares commonalities with Clark's typology of depictions: Indexed depictions would be instantiations of cataphoric integration; the first kind of integration by positioning ("through occupying a syntactic gap") would cover embedded and independent depictions; the second kind of integration by positioning ("through temporal overlap") would include adjunct depictions.

Ladewig (2020) goes a step further and looks into "interrupted utterances," that is utterances with an empty slot at the utterance-final position. With experiments, she shows that manual gestures can, much like canonical verbal constituents, be used to fill the empty slots in interrupted utterances and become an integrated part thereof, both syntactically and semantically. Coming from a different tradition but equally notable is the research conducted by Keevallik (2010, 2015, 2017, 2018, 2020), who systematically explores "bodily quoting," a phenomenon in the context of dance instruction where the instructor employs bodily movements where a verbal quotation would normally be, in order to demonstrate the contrast between correct and incorrect performances to the students. Focusing on sequential temporality and drawing on data of multimodal interaction in multiple languages, she further demonstrates how verbal elements and bodily actions are mutually adapted in real time to create emergent multimodal patterns.

Analogous findings have also been reported from interactions in communicative ecologies other than those between hearing speakers of spoken languages. In the field of sign linguistics, for instance, Ferrara, Hodge, and Johnston have observed that enactments can be sequentially integrated into Auslan (Australian Sign Language), where the enactments can function in place of fully lexicalized manual signs, filling syntactic gaps as well as inferring or expressing semantic relations (Ferrara and Johnston, 2014; Hodge and Johnston, 2014). Based on her fieldwork on the alternate sign languages (see Kendon, 1988b) in the Arandic speaking communities of Central Australia, Green (2014) investigates how manual signs can be employed in discourse, with or without co-occurring speech, depending on the social protocols applicable to the current discourse. Specifically, she shows that signs can, in the absence of simultaneous speech, replace spoken lexical items in utterances, in some instances creating multimodal composite utterances with semantic contributions from both speech and sign (see also Green and Wilkins, 2014).

Finally, recent years have seen attempts at incorporating gesture into the theoretical framework of linguistic analysis, coming from various theoretical orientations and with different approaches [e.g., "integrated message model" (Bavelas and Chovil, 2000); "composite signal" (Clark, 1996); "composite utterance" (Enfield, 2009); "multimodal grammar" (Fricke, 2012); multimodal negation (Harrison, 2018); incorporation of gesture into Cognitive Grammar (Kok and Cienki, 2016); "mixed syntax" (Slama-Cazacu, 1976)]. Construction Grammar, in particular, sees a recent debate on Multimodal Construction Grammar (e.g., Steen and Turner, 2013; Schoonjans et al., 2015; Cienki, 2017; Hoffmann, 2017; Schoonjans, 2017; Ziem, 2017; Zima and Bergs, 2017). Arguing for nonverbal signals being as integral to language as canonical speech, these studies touch upon cases of gestures without simultaneous speech, acknowledging their crucial role in language use, but the primary focus remains on gesture-speech co-occurrence.

Essentially, phenomena on the right half of the continuum exemplify prototypical cases of "marginalia," which, as Dingemanse (2017b, p. 195) identifies, are "typologically unexceptional phenomena that many linguists think can be ignored without harm to linguistic inquiry" — though not rare, "linguistic practice assigns them to the margin by consensus." The handful of existing studies above only provide a first glance at, or around, the largely overlooked domain that is iconic gestures without co-occurring speech, revealing how limited our current understanding still is. Indeed, while certain subgroups of such cases have been studied, there has yet to be a general, systematic survey of the phenomenon itself — one that delimits it, explores its relations to speech, and examines how such gestures contribute to the resulting multimodal discourse — not least in spoken language interactions. In the following, we take a first step in this direction, within Clark's framework of depicting, as it offers a schematic perspective on iconic meaning communication in general.

SPEECH-EMBEDDED NONVERBAL DEPICTIONS

Up to this point, we have been arguing for the relevance of the overlooked domain from the theoretical side, contextualizing it against relevant research. In this section, we turn our attention to the empirical side of the phenomenon, which we now zoom in and elaborate on as "speech-embedded nonverbal depictions." In addition to a detailed delimitation of the phenomenon based on real-life examples from our corpus, we also present a preliminary sketch of the complexity exhibited by such depictions.

While "speech-embedded nonverbal depictions" is not an opaque term, in order to properly identify our target phenomenon in relation to existing studies bordering the overlooked domain in the literature, we further define such depictions in more technical terms, as

— depictions that are embedded in speech, but that are not depictions of non-depictive speech,

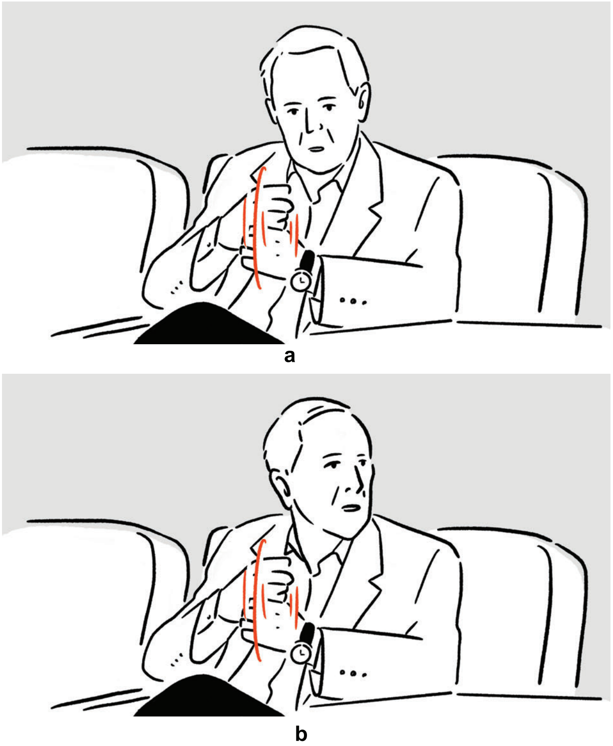


FIGURE 6 | Depictions in (9).

where depictions are understood in the sense defined in Clark's (2016) staging theory. The following excerpts present prototypical cases of such depictions.

(9) Bob Newhart on getting feedback from the audience when performing in the rain: "This one umbrella starts to [*stacks right fist on top of left fist in center-center, as if holding an umbrella, lightly shaking both arms vertically*][a], starts to [*stacks right fist on top of left fist in center-center, as if holding an umbrella, lightly shaking both arms vertically*][b], starts to jiggle."

— *Conan*

(10) Zooey Deschanel on being refused priority boarding when traveling with her baby daughter: "and I was like, but [*moves both arms back and forth parallel to frontal plane, elbows bent, both palms up, left palm placed on top of right palm*][a]. She needs to go on the plane."

— *The Ellen DeGeneres Show*

In (9), Newhart recounts his experience of doing stand-up comedy in open air, where some of the audience were holding an umbrella because of the rain, and where, at some point, one umbrella started to jiggle because the person holding it was laughing. In the temporal "gaps" in his speech, he depicts, using mainly movements of the hands, arms, and shoulders, the jiggling of one of the umbrellas, thereby communicating the original scene of the event in an iconic way, with fine-grained motoric details. Sharing her experience of being denied priority boarding even though she was traveling with her baby daughter, Deschanel depicts in (10), after the word *but*, her reaction upon being so told, displaying actions typically associated with holding and rocking an infant, thereby enacting the scene, with imagistic details, to the audience of the talk show. In both (9) and (10), the nonverbal depictions are embedded in speech, filling the temporal "gap" therein. Employed to communicate meaning without the support of temporally overlapping speech, they exemplify the canonical use of speech-embedded nonverbal depictions in interaction.

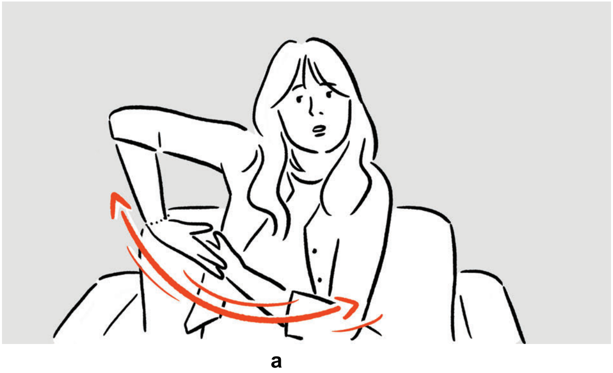


FIGURE 7 | Depiction in (10).

It needs to be reiterated that speech itself can be depictive, descriptive, and indicative, and that cases of depictive speech fall under the category of depiction as well. For instance, the words in (11), which directly precede (10), are depictive of words uttered in the past, therefore a depiction.

(11) Zooey Deschanel on being refused priority boarding when traveling with her baby daughter: "and they were like, no, like, the people who get on first pay a lot of money for this privilege."

— *The Ellen DeGeneres Show*

Quoting the ground crew member who denied her priority boarding, Deschanel is effectively staging a depiction of a past event, except that the past event is one where descriptive speech is uttered.

Ubiquitous and complex in their own right, depictions of descriptive speech — that is, canonical quotations — have long intrigued linguists and have a rich and extensive literature (e.g., McGregor, 1997; Tannen, 2007; Vandelanotte, 2009; Buchstaller, 2014; Spronck and Nikitina, 2019; see also Hodge and Cormier, 2019 for discussion in relation to depicting). Though indeed frequently observed in our corpus, such tokens are not included in our analysis, where the aim is to draw attention to overlooked phenomena in the literature. While eventually consolidating depictions across all modalities and signaling methods would be optimal, at the current stage, excluding canonical quotations allows us to prioritize focus on depictions that have hitherto eluded the attention of researchers — that is, depictions that are really marginalia (Dingemanse, 2017b) in the literature. This is reflected in our technical definition of speech-embedded nonverbal depictions presented above, where depictions of non-depictive speech are excluded, allowing us to focus on the core cases of iconic meaning communication.

Importantly, the exclusion of descriptive and indicative speech does not rule out cases of depictive speech from our analysis. A broad concept itself, depictive speech subsumes a number of phenomena and has been given various labels, such as multimodal quotation, sound symbolism, interjection, onomatopoeia, and ideophone (see e.g., Kita, 1997; Dingemanse, 2013, 2015), many of which have only recently been picked up in the cognitive-functional linguistics literature. Not only do they call for fuller exploration, they are also curious from the perspective of depicting and multimodality, as creative multimodal strategies are usually needed to establish iconic mappings between depictive speech and its depicted scene. Indeed, building on Dingemanse's (2013) study, Clark (2019) identifies ideophones as depictions in the verbal modality, distinguishing between "free" and "codified" depictions, which can be illustrated, respectively, by the following examples, taken from our corpus.

(12) Jennifer Garner on accidentally kayaking into a busy harbor: "There were like [*vocalizes* brrr *sound; moves both hands slowly from left to right, palms facing each other, fingers spread, distance between palms constant*][a], like big boats."

— *The Tonight Show Starring Jimmy Fallon*

(13) Chris Evans on bullying his brother (Scott Evans, seated to his left), in childhood: "And I just had the book, and just, [*vocalizes* whack, *moves left hand in a curve, from lower right periphery to upper left extreme periphery, close to where Scott's head is*][a], and I hit him."

— *Late Night with Seth Meyers*

In (12), where Garner recalls encountering big boats as she accidentally kayaked into a busy harbor, the big boats are referred to iconically. This is done, not just by her highly metonymic "bounding" (Streeck, 2008) manual gesture — where, drawing on the contiguity relation between her hands and the depicted object (Mittelberg and Waugh, 2014), the empty space between her hands is mapped onto some generic big boat — but also by the low-frequency *brrr* sound, depictive of the sound of boat horn. "Created *de novo*" (Clark, 2019, p. 235), *brrr* instantiates a free depiction. In (13), Chris Evans recounts hitting his brother with a thick book, depicting the scene by deploying a set of manual gestures, and, on top of that, *whack*, which is an ideophone codified in the English vocabulary for the sound of heavy strikes, and which is thus a codified depiction. Importantly, in both of the cases, the speaker establishes physical, specifically auditory, resemblance between the depictive speech and the depicted sound creatively, as it is humanly impossible to literally reproduce the latter.

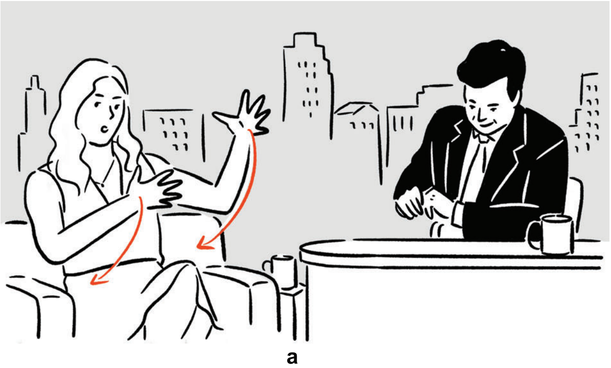


FIGURE 8 | Depiction in (12).

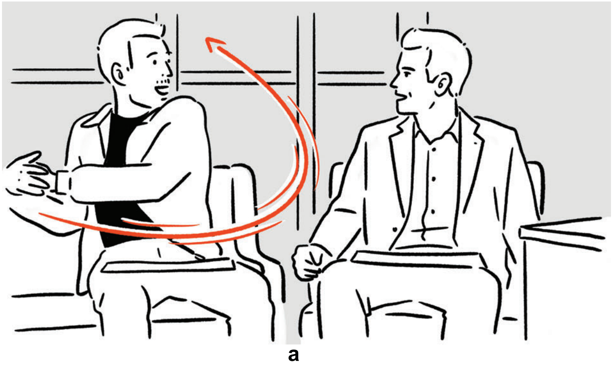


FIGURE 9 | Depiction in (13).

It is following the definition spelled out in the present section, and with the modality-agnostic gesture phrase as the basic depiction unit, that the 217 tokens of speech-embedded nonverbal depictions were identified in our American TV talk show corpus. In what follows, we present further theoretical and methodological considerations — pertaining to the issue of embedding in particular — resulting from a closer examination of the 217 target tokens, as well as some observations regarding the internal complexities of the depictions.

Embedding

In addition to distinguishing speech with different semiotic functions, another key notion that needs clear delimitation is embedding. It is a term that is particularly tricky because it can be understood either in terms of function or form, which are often conflated.

Clark (2016, pp. 325–326), in his typology, makes the functional distinction between embedded and independent depictions, with the former functioning as "parts of utterances" and the latter making "independent contributions to the discourse." Empirically, this distinction is easily blurred. Consider the depiction in (14).

(14) Conan O'Brien: "How do you do that, do that again?"
Kristin Chenoweth: "[*sings syllables* aye-ah *in high pitches*]"

— *Conan*

With the guest having just demonstrated some high-pitched singing, the host, impressed, asks the guest how that is done and requests that she do it again. In response to this, Chenoweth simply depicts her own singing, rather than verbally describe her singing technique. Contributing to the discourse without adjacent or co-occurring speech, Chenoweth's depiction exemplifies what Clark (2016) identifies as an independent depiction.

Viewed on a more schematic level, the category boundary becomes less clear-cut. Among other things, the guest's depiction only makes sense with the preceding discourse considered; it is co-dependent with the host's question in carrying out their global function as a question-and-answer pair. As is the case for any signal in language use, the contributions made by independent depictions to the discourse are seldom, if ever, truly independent, a fact that undermines the functional basis of the embedded-independent distinction. In this sense, independent depictions are really as embedded as embedded depictions, except not on the level of the word or phrase, but on the level of the sequential organization of the interaction. Both types of depictions function as if they were verbal constituents, contributing meaning iconically without simultaneously co-occurring speech.

From a form-based perspective, embedding can be understood in temporal terms, that is the temporal overlap between a depiction and its adjacent speech. In discussing the temporal placement of depictions, Clark (2019, p. 241) points out that both embedded and independent depictions are "slotted" into utterances "without breaks or overlap,"[12] filling temporal gaps in utterances. That is, embedded and independent depictions do not differ in this regard. While there might be operationalizable ways to systematically untangle the overlap between embedded and independent depictions in function [see e.g., Lehmann's (1988) gradient approach to clause linkage along multiple continua; and Hodge (2014) on clause-like units in signed language], they simply exhibit no difference in form as far as temporal overlap is concerned.

In accordance with our annotation, we adopt a form-based definition of embedding, in temporal terms, which effectively dissolves the categorical distinction between embedded and independent depictions in Clark's typology, rendering both as instantiations of embedded depictions in our corpus. Specifically, we define an embedded depiction as one whose stroke phase does not overlap with temporally co-occurring speech — as per our definition of the depiction unit, the stroke phase of a depiction is to be understood in the broad, modality-agnostic sense, as a schematization from the stroke phase of a manual gesture, and refers to the central, meaningful part of the movement of a depiction. In addition to allowing us to focus on the core component of depictions, this criterion also yields a more accurate picture of depiction embedding: As is the case for manual gestures, speakers in our corpus, in employing embedded depictions, are often observed preparing themselves ahead of the slot, timing the stroke of the depiction to be executed within the precise time frame of the slot.[13]

Reconsider example (13), repeated here as (15).

(15) Chris Evans on bullying his brother (Scott Evans, seated to his left), in childhood: "And I just had the book, and just, [*vocalizes* whack*, moves left hand in a curve, from lower right periphery to upper left extreme periphery, close to where Scott's head is*], and I hit him."

— *Late Night with Seth Meyers*

Recalling how he left a scar on the forehead of his brother by hitting him with a thick paperback book, Chris Evans stages a depiction after the second *just*, utilizing both his entire left arm and the codified ideophone *whack*. To demonstrate the full extent of the whacking, the speaker can be seen already retracting his left arm to his right at the second *just*, and retaining a gesture hold until after the word *him*. The gesture phrase therefore spans from *just* to after *him*. Despite the temporal overlap between speech and some of the depiction phases, we view the depiction in (15) as embedded, since its stroke is timed to fill the empty "slot" in the speech, in a sequential and not simultaneous manner, without temporal overlap.

Likewise, reconsider the jiggling example in (9), repeated here as (16).

(16) Bob Newhart on getting feedback from the audience when performing in the rain: "This one umbrella starts to [*stacks right fist on top of left fist in center-center, as if holding an umbrella, lightly shaking both arms vertically*], starts to [*stacks right fist on top of left fist in center-center, as if holding an umbrella, lightly shaking both arms vertically*], starts to jiggle."

— *Conan*

Following the segmentation established above, this excerpt contains two depiction phrases, therefore two tokens of depictions. In addition to the preparation before the first depiction and the hold after the second, a "depiction hold" is also observed between the two depictions. In fact, all of the words included in the excerpt overlap temporally with some depiction phase. The two depictions are nevertheless embedded depictions, as their stroke phase does not temporally coincide with speech, but takes up a temporal gap in the sequence of the embedding speech.

In addition to preventing form-function conflation, defining embedding in temporal terms also helps to avoid some of the problems resulting from underspecification, such as those enumerated about (8), which lies on the boundaries of adjunct, embedded, and independent depictions in Clark's typology. It is repeated here as (17).

(17) Tracy Morgan on the quality of his facial muscles: "Yeah, I'm your rubber-band man, [*vocalizes* brbrbrbr *sound,*

[12] The criterion "without breaks or overlap" also proves problematic empirically. See discussion on temporal overlap immediately below.

[13] Clark in his (2016, p. 340) paper touches upon what he calls "phases of discourse" and "discourse timing," but does not explicitly elaborate on the different ways in which these notions relate to one another in the four depiction types he identifies.

shakes head sideways quickly, causing facial muscles to vibrate accordingly]."

— *Conan*

Despite the functional affiliation between the depiction and *rubber-band man*, since the stroke of the depiction takes place only after *man*, it is annotated as an embedded depiction in our corpus. Indeed, while there is no definitive way of determining whether the depiction functions more like an adjunct or a separate utterance, it is objectively, temporally embedded in the discourse.

One final implication of defining embedding in temporal terms pertains to the intersection between depictions and verbal indices. As mentioned, Clark's (2016) definition of indexed depictions — as those that are indexed by indexical expressions in speech, such as *this* and *there* — does not explicitly address the fact that there exist two distinct kinds of verbal indices. Consider the indexical devices in (18), where the host claims he is not quick-witted enough to be on a game show, and (19), where the guest demonstrates her peculiar way of nodding.

(18) Conan O'Brien on his lack of quick wit: "I don't have that, quick, [*snaps fingers of left hand thrice*][a]."

— *Conan*

(19) Emily Blunt on her impassive backchannels: "I just go like this [*nods head repetitively, quickly, but with little movement; maintains gaze at Seth Meyers, seated to her left*][a]."

— *Late Night with Seth Meyers*

Although verbal indices are present in both excerpts, they function in distinct manners. In (18) — where Conan depicts quick wit with finger snapping, which is associated with moments of epiphany, and therefore, metonymically, with people adept at witty comebacks — the verbal demonstrative *that* modifies what follows it, making it an indexical modifier. Accordingly, it is not that *that* indexes *quick, [*snaps fingers*]*, but that *that, quick, [*snaps fingers*]*, as a whole, indexes the kind of quick wit the host is referring to. In contrast, the indexical *this* in (19) — where Blunt simply depicts her own nodding — is a pronoun with indexical functions. As such, *this* by itself directly indexes the speaker's peculiar head nod. In other words, where a depiction is employed in connection to an indexical modifier, what is indexed is not the depiction; where an indexical pronoun is used in conjunction with a depiction, it is the depiction that is indexed. Consequently, if indexed depictions are those that are indexed by indexical speech, they should only include cases of indexical pronouns, as in (19), and not cases of indexical modifiers, such as (18) — since in the latter, it is not the depiction, but the combination of the verbal index and the depiction, that is indexical.

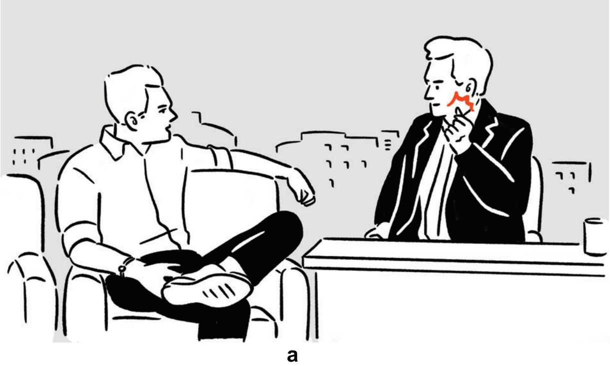


FIGURE 10 | Depiction in (18).

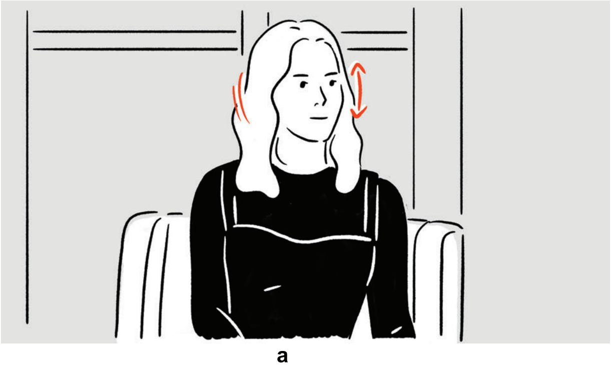


FIGURE 11 | Depiction in (19).

Importantly, while the distinction between indexical modifiers and pronouns is crucial, the relations between verbal indices and depictions — specifically whether a depiction is indexed by a verbal index — are a functional concern. In other words, indexation is an issue on a dimension independent of our form-based definition of temporal embedding: A depiction can be indexed, embedded, neither, or both. Since both of the depictions in (18) and (19) occupy a temporal gap in speech without their stroke overlapping with speech, they are categorized as embedded depictions in our corpus, regardless of the fact that one of them is also verbally indexed and the other not. Though crucial to the understanding of depictions in general, a full-fledged exploration of the relations between depictions and verbal indices is left for further research.

Complexities of Speech-Embedded Nonverbal Depictions

In addition to the issues concerning embedding alone, the tokens of speech-embedded nonverbal depictions in our corpus also exhibit complexities in other regards, of which we now offer a brief sketch. While the primary aim of the present paper is to draw attention to the phenomenon of speech-embedded nonverbal depictions, rather than to present an in-depth analysis, the following offers a glimpse into their theoretical and empirical potential, which further underscores the need for research on this overlooked topic.

Consider the depictions in (20), where Mulaney recalls stumbling wearing high heels, and where the depiction is preceded by *like*.

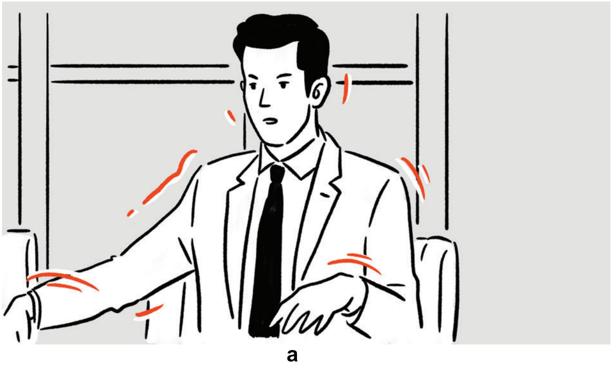


FIGURE 12 | Depiction in (20).

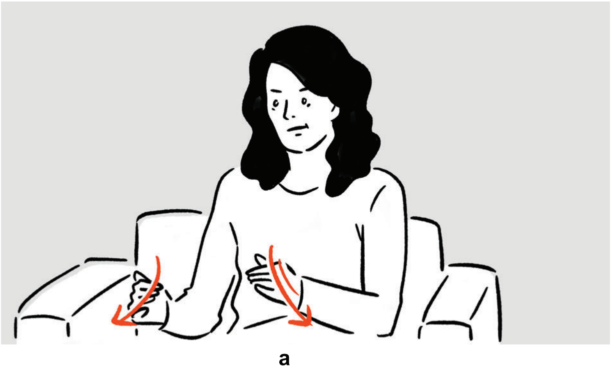


FIGURE 13 | Depiction in (21).

(20) John Mulaney on walking on heels: "It's like, it was like, [*stretches out both arms sideways, tilts torso in different directions, as if trying to find balance*]^a, when a, when a cow's born."

— *Late Night with Seth Meyers*

Perhaps unsurprisingly, tokens preceded by *like* are frequently observed (see Golato, 2000; Streeck, 2002), due to *like*'s function as a quotative (see e.g., Tagliamonte and Hudson, 1999; Macaulay, 2001; Vandelanotte and Davidse, 2009), which indexes many of the functions depictions serve, such as quotation, enactment, demonstration, and pantomime (see Hodge and Cormier, 2019). What makes things less straightforward, however, is that *like* also often functions as a marker signaling hesitation or hedging, among many other things (see Miller and Weinert, 1995; D'Arcy, 2017). It further complicates the picture that, in cases like (20), the depiction is sometimes followed by verbal elements whose meaning overlaps with that of the depiction.[14]

With the stumbling depiction preceding *when a cow's born*, it is plausible the depiction in (20) is the physical manifestation of the speaker's thought process, specifically the mental simulation of the action he is trying to verbalize, which eventually results in *when a cow's born* (e.g., de Ruiter, 2000; Hostetter and Alibali, 2008; see also Streeck, 2009 on "ceiving"). At the same time, it is also not unreasonable to suspect the depiction serves as a filler, one that fills the uncomfortable pause resulting from the speaker's word search (see Goodwin and Goodwin, 1986; Gullberg, 1998; Hadar and Butterworth, 1997; Navarretta, 2015), before the speaker is able to "find their words."

However, we also repeatedly come across cases like (21).

(21) Tina Fey on doing serious choreography: "If I had to be on Dancing with the Stars, I would be so shark-eyes,[15] I would be like [*gazes at the front, into the distance; moves both arms in parallel, elbows bent, as if rowing a boat*]^a, I would so panic all the time."

— *The Ellen DeGeneres Show*

Here Fey depicts how awkward and uncoordinated her movements would be if she were ever to go on a dance show. Like (20), the depiction in (21) is also preceded by *like*. Unlike (20), the depiction in (21) does not precede, but rather follows, *shark-eyes*, the verbal element whose meaning is similar to that of the depiction. In other words, the speaker first communicates the meaning verbally, saying *shark-eyes*, before staging a nonverbal depiction with highly similar semantics. The fact that the speaker first communicates her idea verbally, and then still proceeds to stage a depiction that is semantically "repetitive" — with the identical speech frame of *I would be* no less — shows that such nonverbal depictions cannot be conveniently dismissed as word-search fillers.

Indeed, cases of "multimodal iteration" (Hsu et al., to appear; cf. Johnston, 1996 on the "spiral" manner in which signing can unfold in Auslan), that is the phenomenon where the speaker communicates meaning in multiple combinations of modality and signaling method — specifically, in (20) and (21), verbal description and gestural depiction — may point to nonverbal depictions having different communicative potentials than descriptive speech (see Mittelberg, 2014 on "mediality effects"). In addition to exhibiting cross-modal dialogic resonance (see Du Bois, 2014), such tokens also showcase the reciprocal framing across modalities, whereby verbal and nonverbal elements profile certain aspects of one another (Kendon, 2004; Ferrara and Hodge, 2018). The mechanisms at work here may in turn contribute to the long-lasting debate whether gesture and speech are two separate processes, or manifestations of one single process (e.g., McNeill, 1992, 2013b; de Ruiter, 2000; Kita, 2000; Kendon, 2004), further adding to the reasons why speech-embedded nonverbal depictions deserve more attention.

Also strengthening the case for speech-embedded nonverbal depictions is the fact that they are observed embedded across different syntactic levels, from the level of the word (e.g., Example 9), phrase (e.g., Example 12), clause (e.g., Example 20), all the way to the level of the discourse (e.g., Example 14). The following depictions further exemplify this versatility.

(22) Lil Rel Howery on texting without looking at the screen: "People are just that good where they can just [*gazes at the*

[14] The observations made in this paragraph about *like* are also largely applicable to *just*, as in (15) and (22).

[15] If someone has shark eyes, it means the person's gaze is empty and absent.

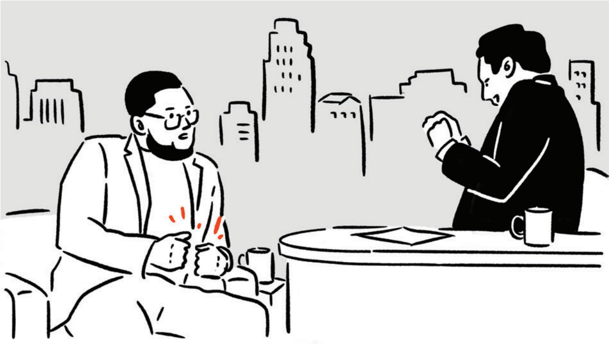


FIGURE 14 | Depiction in (22).

front, into the distance; places both fists above lap, at lower center, flipping both thumbs up and down quickly, as if typing on a phone][a]."

— *The Tonight Show Starring Jimmy Fallon*

(23) Cardi B on being mischievous as a kid: "I was like, ok I know, [*points with right index finger stretched, fingertip moving from center-center to right extreme periphery along a straight line*][a], go to the principal's office."

— *The Ellen DeGeneres Show*

In (22), Howery expresses his frustration with people who type on their phone without looking at the screen. In this case, the depiction is embedded on the level where a complex verbal phrase would otherwise be embedded. In (23), Cardi B stages how, after some mischief, her teacher asked her to go to the principal's office. Here the embedding takes place on the level of the sentence, the depiction functioning like an imperative sentence otherwise would.

Though further research is needed, the versatility in syntagmatic depiction-speech integration already suggests the capability of nonverbal signals in "substituting" for structurally diverse verbal constituents, both in form and function. This echoes Ladewig's (2020) recent findings, potentially also lending support to the view that nonverbal depictions as form-function pairings are not unlike verbal constituents — at least in the sense of Construction Grammar (Croft, 2001) and Cognitive Grammar (Langacker, 2008). This, of course, warrants a separate discussion that is beyond the scope of the present paper (see Ferrara, 2012; Hodge, 2014; Kok and Cienki, 2016; Wilcox and Occhino, 2016; Ruth-Hirrel and Wilcox, 2018; Ladewig, 2020).

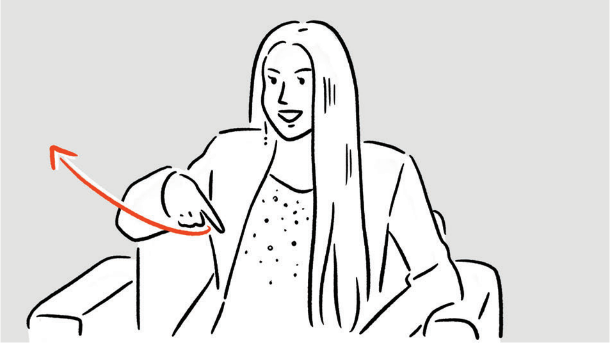


FIGURE 15 | Depiction in (23).

The complexity and full potential of speech-embedded nonverbal depictions are also evident paradigmatically. For instance, reconsider once again the depictions in (9), repeated here as (24).

(24) Bob Newhart on getting feedback from the audience when performing in the rain: "This one umbrella starts to [*stacks right fist on top of left fist in center-center, as if holding an umbrella, lightly shaking both arms vertically*], starts to [*stacks right fist on top of left fist in center-center, as if holding an umbrella, lightly shaking both arms vertically*], starts to jiggle."

— *Conan*

Here Newhart says an umbrella starts to jiggle, but what he depicts in the two embedded depictions is in fact not the jiggling of the umbrella per se, but the cause of the jiggling, namely the person laughing whilst holding the umbrella, who is in turn represented by Newhart's fists. Despite the "mismatch," Newhart is able to get his message across because of the metonymic relations that are at play here: part for whole (the fists for the umbrella holder), and cause for effect (the umbrella holder's action for the umbrella's movement). Furthermore, the phrase *this one umbrella starts to jiggle* (whether the notion of jiggle is communicated through the depiction or the word *jiggle*) is itself a metonymic way of saying a member of the audience starts to laugh (effect for cause: the umbrella's movement for the person's action).

Paradigmatic complexities are also manifest in the observation that speech-embedded nonverbal depictions are sometimes employed back to back, such as in (25), in which the host demonstrates how he would not be able to refrain from actually eating if he were to play a role that requires eating on scene.

(25) Conan O'Brien on being unable to refrain from savoring the food if required to eat on scene: "I'd be, even in a drama, they'd be like, Conan's the murderer, [*vocalizes* kahm-ahm, *moves mouth as if biting and chewing; moves hands in parallel, from lower center toward upper center near own mouth, fingers touching on both hands, as if holding a hamburger*][a] — [*vocalizes* hum-um, *sucks own fingers*][b] — [*stretches out right index finger in upper right periphery, as if signaling some imaginary addressee to wait until he is done eating; moves mouth as if chewing*][c]."

— *Conan*

In the first set of actions, Conan depicts himself ferociously munching on some burger type of food item; the second depiction includes the finger sucking action typically associated with someone enjoying fast food; in the third and final set of actions, Conan depicts how he would prioritize actually eating over acting. Notably, in all three of the depictions, which are

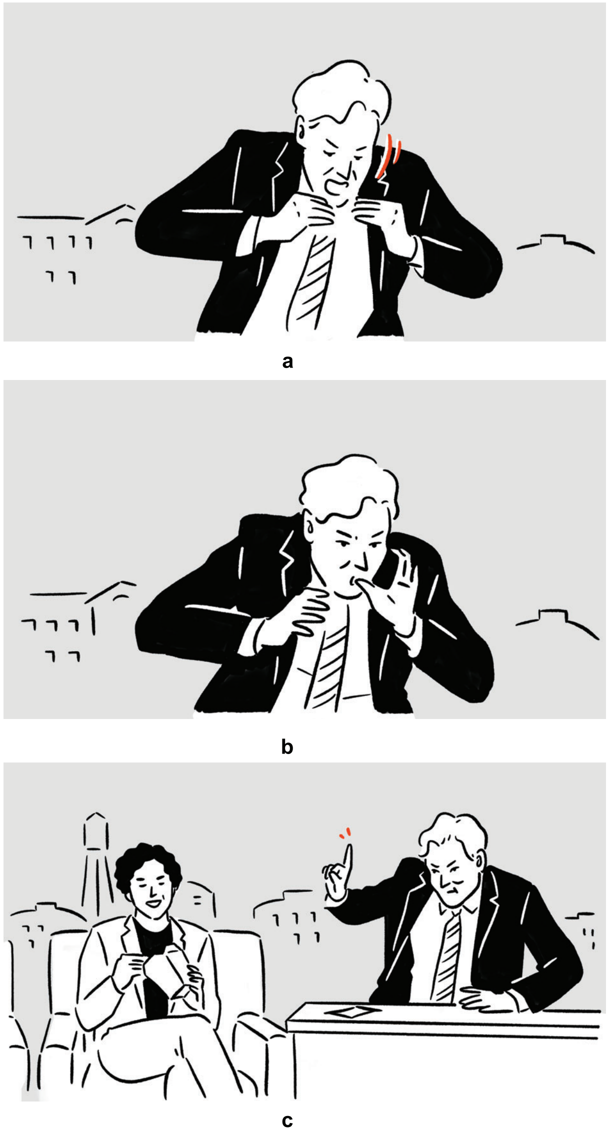


FIGURE 16 | Depictions in (25).

staged consecutively without speech "intervening," Conan can be observed maintaining the same bodily posture, consisting primarily of raised shoulders and upper arms.

What really sets this example apart is the fact that the understanding of the later depictions hinges on the understanding of the earlier ones. Without the first depiction, the finger-sucking gesture in the second depiction would be a lot harder to make sense of. Likewise, were the first two depictions absent, the third depiction would hardly be decipherable on its own. In other words, the later depictions build and elaborate on prior depictions, along the same storyline. Together, the co-dependent depictions, bound together by the common thread that is Conan's sustained posture, contribute to a composite structure with a complex meaning.

On a more theoretical level, complex composite depictions (Hsu, 2019) like (25) are significant to the discussion on the role of nonverbal signals in language, and therefore also to the above-mentioned Multimodal Construction Grammar debate, as they demonstrate that even singular actions (Müller, 2010, cited in Ladewig, 2014) — that is, actions created and assembled *ad hoc* (see Brône and Zima, 2014), for highly local purposes — can be combined to create larger structures, undermining the argument that gestures are not "linguistic" simply because of their lack of recurrence and low frequencies (see Schoonjans, 2017). The observation that the component depictions in the composite series share a common posture as their "base" (Hsu, 2019), also echoes comparable phenomena that have been identified in the literature, such as "locution cluster" (Kendon, 1972), "catchment" (McNeill, 2005), and "frame hold" (Sowa, 2006).

Cases of composite depictions can be further complicated by viewpoint changes. Consider again the depiction sequence in (6), repeated here as (26).

(26) Lauren Ambrose on backstage costume change on Broadway: "I mean sometimes it's like twenty seconds, for like, full-on, [*vocalizes whistle-like* fsss *sound, moves both hands vertically, fingers spread, in opposite directions, in front of head and torso*] — [*vocalizes whistle-like* ffft *sound, gazes at the front, into the distance, moves both hands along sagittal axis away from body, fingers spread, palms away from body*]."

— *Late Night with Seth Meyers*

Similar to (25), the two depictions in (26) depict two subevents unfolding in sequence which are part of a larger event: The backstage staff on Broadway first changed Ambrose's makeup and costume, and, after that, pushed her back to the stage. In addition to the composite structure, a striking viewpoint shift is observed between the two depictions. In the first depiction, the speaker takes on her own viewpoint in the depicted event (one can also argue that, since her hands depict the staff members' hands, she also takes on the staff members' perspective simultaneously; see e.g., McNeill, 1992; Parrill, 2009; Dancygier and Sweetser, 2012 on dual viewpoint). In the second, she takes on the perspective of the backstage staff member who pushed her back to the stage. Remarkably, the only overt cue signaling this shift in perspective is her gaze behavior: During the first depiction, the speaker appears to be looking at the host; during the second, her gaze is averted, focused instead on something in the distance. Tokens such as this echo findings in recent studies (e.g., Sidnell, 2006; Sweetser and Stec, 2016), which situate speech-embedded nonverbal depictions at the intersection between gesture, viewpoint, and gaze (see also Stec et al., 2016; Janzen, 2017).

The communicative potential of speech-embedded nonverbal depictions can also be exploited jointly across multiple speakers, as is the case in (27), an extended excerpt of which (13) is a part. As Chris Evans recounts hitting Scott Evans, his brother, with a thick book, Scott, seated to Chris's left, joins in the storytelling, using not words, but depictions.[16]

[16] Given the complexity resulting from the temporal overlap between the two speakers, and informed by the conventions of Conversation Analysis, we employ

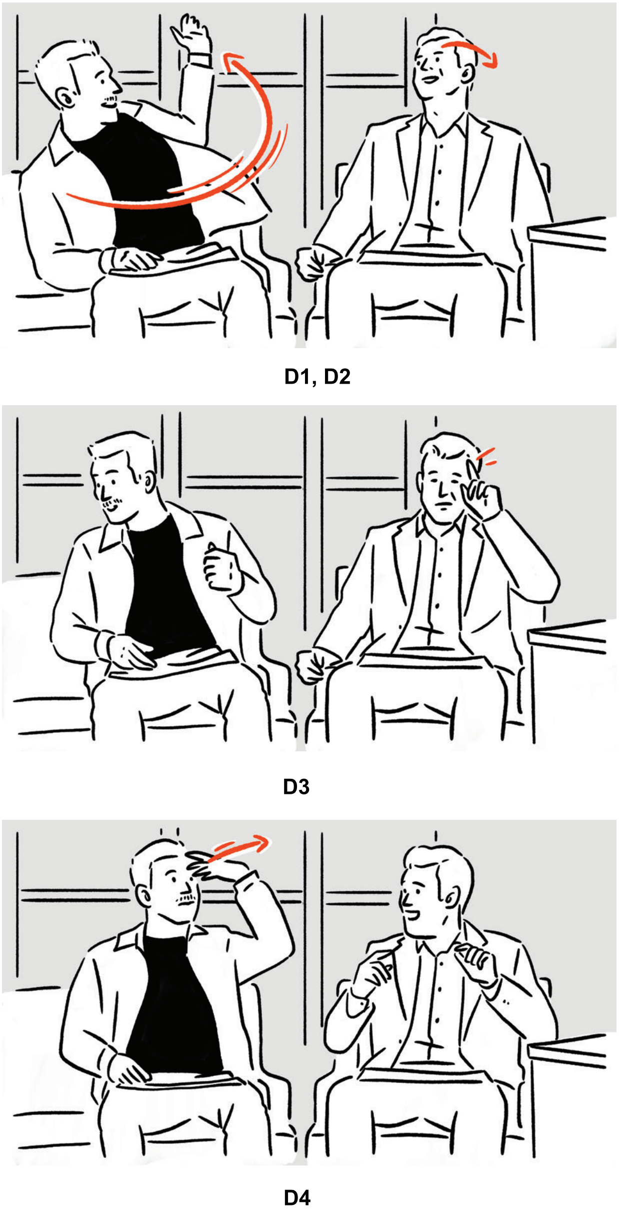


FIGURE 17 | Depictions in (27).

(27) Chris Evans (A) on bullying his brother, Scott Evans (B), who is seated to his left, in childhood:

A: And I just had the book,
   and just, *[*D1*]*, and I hit him,
B: *[*D2*]*
A: and as *soon as I* hit him, [*D4*]
B: *[*D3*]   *

D1: Vocalizes *whack*; moves left hand in a curve, from lower right periphery to upper left extreme periphery, close to where B's head is.
D2: Tilts head away from A.
D3: Traces scar on left forehead with left index.
D4: Vocalizes *brrr*; touches forehead with fingertips of left hand, fingers touching, moves left hand toward upper left extreme periphery, spreading fingers along movement.

— *Late Night with Seth Meyers*

Almost as soon as Chris stages the *whack* depiction, Scott is seen staging the second depiction, which is an enactment of his response to being hit by Chris. When Chris is at *soon as I*, Scott again contributes to the story, depicting the scar by locating and finger-tracing its shape on his forehead, before Chris stages the fourth depiction, which demonstrates, in an exaggerated manner, the spurting of the blood that came out of Scott's forehead. The series of depictions, from both parties, goes on beyond the excerpt. As in (25) and (26), the depictions are co-dependent in meaning. Unlike in (25) and (26), the depictions in (27) are not all staged by one single speaker, but are staged jointly by two speakers, with causation between the depictions, bringing in the complexities of an additional, interactional dimension to the analysis.

The above is a very brief sketch of some of the complexities of speech-embedded nonverbal depictions, based only on tokens taken from our American TV talk show corpus, where the annotated data still await in-depth analysis. The rich and challenging cases this alone has already provided us with, nonetheless hint at the fact that speech-embedded depictions are not merely theoretically significant, but abundant in curiosities of language use and interaction as well.

CONCLUDING REMARKS

Drawn to nonverbal iconic language use, and informed by Clark's recent account of depicting in everyday interaction, we turned to video recordings of American TV talk shows, a context rich in depictions. The annotation of the data proved less straightforward than expected, an issue that underlines our limited understanding of this domain of research. In addition to operationalizing relevant theoretical notions, a critical reconsideration of depiction-speech relations, on the basis of Clark's typology of depictions, was carried out, resulting in a gradient reconceptualization of depictions in terms of meaning contribution from non-depictive speech and depictive signals. This led to the identification of a largely overlooked domain — cases where meaning is communicated through iconic nonverbal signals, without temporally co-occurring speech — which we zoomed in on as "speech-embedded nonverbal depictions." Taking into consideration existing literature as well as the variety of tokens in our corpus, we arrived at a carefully delimited definition of such depictions, in turn bringing to the fore numerous observations, many of which pertain to

asterisks (rather than brackets, which in the present paper already indicate nonverbal signals without co-occurring speech) to mark the beginning and end of simultaneous events (cf. Mondada, 2016). To facilitate readability, the depictions are dubbed "Dn" and described after the text excerpt.

current discussions in cognitive linguistics, gesture studies, and multimodal communication.

As an initial step into the largely uncharted territory, it goes without saying the present study is limited in several ways. Among them is the type of data examined. The majority of the speakers in our corpus are professional actors or comedians, a fact that likely has an effect on the frequency, elaborateness, and spontaneity of the depictions they stage. Nonetheless, while true spontaneity is hardly attainable, it is undeniable that American TV talk shows, which are themselves a specialized context, contain unscripted elements. In addition, as the staging theory (Clark, 2016) suggests, performativeness is an inherent aspect of depicting (as has also been reported for Auslan; see Hodge and Ferrara, 2014), that is the signaling of meaning through showing. Dramatizations and exaggerations have also been reported to be common in narratives in general (see e.g., Bavelas et al., 2014; Stec et al., 2015). On a more schematic level, the present paper is focused primarily on spoken language interactions, due in part to the fact that the topic of the current study is particularly marginalized in spoken language linguistics. This is in contrast with signed language linguistics, which sees many relevant phenomena being more established topics in its literature (see among many others the above-mentioned Liddell, 2003; Wilcox and Occhino, 2016; Ferrara and Hodge, 2018). Future studies on the topic will benefit from larger datasets that are more diverse in terms of communicative ecologies (see Beukeleers and Hsu, 2019 for an initial attempt), which will also facilitate quantitative analysis, potentially bringing in insights from a different angle.

The scope of the data notwithstanding, the tokens in our corpus already shed light on some of the natural next steps for depiction researchers to embark on. Among them are depiction-speech relations, multimodal iteration, complex composite depictions, viewpoint in multimodal interaction, as well as jointly staged depictions and the causation therein. These are the tracks along which we are currently carrying out analysis of the tokens. Though not explicitly touched upon in the present paper, the tokens also point to a number of other directions in which future studies can proceed, such as issues pertaining to intersubjectivity, the performative aspect of depictions, depicting and language acquisition, and motivations for employing speech-embedded nonverbal depictions.

Speech-embedded nonverbal depictions are situated, not only at the crossroads of numerous research traditions, but also among intertwined modalities and signaling methods, which prove tricky to untangle. Nonetheless, as showcased by the versatile and complex ways in which speech-embedded nonverbal depictions are employed in real-life interaction, a full picture of language use will not be complete without a systematic account of such depictions.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

H-CH conceptualized the study, compiled and annotated the data, carried out the analysis, and wrote the manuscript. GB and KF provided guidance in the conceptualization of the study, assisted in the analysis, and contributed to the revision of earlier versions of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The study reported in the present paper is supported by the GSSA scholarship awarded to H-CH by the Ministry of Education of Taiwan.

ACKNOWLEDGMENTS

The authors are grateful to the reviewers for their thorough reading and insightful comments, and to the editor for her assistance along the process. We are also much indebted to Yuga Huang (yugagagahuang.myportfolio.com) for her drawings of the depictions.

REFERENCES

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28, 105–124. doi: 10.1017/S0140525X05000038

Arnheim, R. (1969). *Visual Thinking*. Berkeley: University of California Press.

Bavelas, J. B., and Chovil, N. (2000). Visible acts of meaning: an integrated message model of language in face-to-face dialogue. *J. Lang. Soc. Psychol.* 19, 163–194. doi: 10.1177/0261927X00019002001

Bavelas, J. B., Gerwing, J., and Healing, S. (2014). Effect of dialogue on demonstrations: direct quotations, facial portrayals, hand gestures, and figurative references. *Discourse Process.* 51, 619–655. doi: 10.1080/0163853X.2014.883730

Beukeleers, I., and Hsu, H.-C. (2019). "Complex composite depictions and their semiotic diversity: evidence from gestures and signs," *Paper Presented at the 6th European and 9th Nordic Symposium on Multimodal Communication*, Leuven.

Bloom, P. (2010). *How Pleasure Works: The New Science of Why We Like What We Like*. New York, NY: Norton.

Bressem, J. (2013). "A linguistic perspective on the notation of form features in gestures," in *Body—Language—Communication*, Vol. 1, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 1079–1098.

Bressem, J. (2014). "Repetitions in gesture," in *Body—Language—Communication*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1641–1649.

Brône, G., and Zima, E. (2014). Towards a dialogic construction grammar: ad hoc routines and resonance activation. *Cogn. Ling.* 25, 457–495. doi: 10.1515/cog-2014-0027

Buchstaller, I. (2014). *Quotatives: New Trends and Sociolinguistic Implications*. Oxford: Wiley-Blackwell.

Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington: Indiana University Press.

Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Res. Lang. Soc. Inter.* 25, 163–194. doi: 10.1080/08351819109389361

Cienki, A. (2017). Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Ling. Vanguard* 3(s1):20160048. doi: 10.1515/lingvan-2016-0048

Cienki, A., and Müller, C. (eds) (2008). *Metaphor and Gesture*. Amsterdam: John Benjamins.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. (2003). "Pointing and placing," in *Pointing: Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Psychology Press), 243–268.

Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026

Clark, H. H. (2019). "Depicting in communication," in *Human Language: From Genes and Brains to Behavior*, ed. P. Hagoort (Cambridge, MA: Massachusetts Institute of Technology Press), 235–247.

Clark, H. H., and Gerrig, R. J. (1990). Quotations as demonstrations. *Language* 66, 764–805. doi: 10.2307/414729

Cormier, K., Quinto-Pozos, D., Sevcikova, Z., and Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: comparing depicting constructions and viewpoint gestures. *Lang. Commun.* 32, 329–348. doi: 10.1016/j.langcom.2012.09.004

Cormier, K., Smith, S., and Sevcikova, Z. (2016). Rethinking constructed action. *Sign Lang. Ling.* 18, 167–204. doi: 10.1075/sll.18.2.01cor

Croft, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.

Dancygier, B., and Sweetser, E. E. (eds) (2012). *Viewpoint in Language: A Multimodal Perspective*. Cambridge: Cambridge University Press.

D'Arcy, A. (2017). *Discourse-Pragmatic Variation in Context: Eight Hundred Years of LIKE*. Amsterdam: John Benjamins.

de Ruiter, J. P. (2000). "The production of gesture and speech," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 284–311. doi: 10.1017/cbo9780511620850.018

Dingemanse, M. (2013). Ideophones and gesture in everyday speech. *Gesture* 13, 143–165. doi: 10.1075/gest.13.2.02din

Dingemanse, M. (2015). Ideophones and reduplication: depiction, description, and the interpretation of repeated talk in discourse. *Stud. Lang.* 39, 946–970. doi: 10.1075/sl.39.4.05din

Dingemanse, M. (2017a). Expressiveness and system integration: on the typology of ideophones, with special reference to Siwu. *STUF Lang. Typol. Univ.* 70, 363–384. doi: 10.1515/stuf-2017-0018

Dingemanse, M. (2017b). "On the margins of language: ideophones, interjections and dependencies in linguistic theory," in *Dependencies in Language*, ed. N. J. Enfield (Berlin: Language Science Press), 195–203.

Dingemanse, M. (2019). "'Ideophone' as a comparative concept,"," in *Iconicity in Language and Literature*, Vol. 16, eds K. Akita and P. Pardeshi (Amsterdam: John Benjamins), 13–33. doi: 10.1075/ill.16.02din

Du Bois, J. W. (2014). Towards a dialogic syntax. *Cogn. Ling.* 25, 359–410. doi: 10.1515/cog-2014-0024

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press.

Ferrara, L. (2012). *The Grammar of Depiction: Exploring Gesture and Language in Australian Sign Language (Auslan)*. Ph.D. dissertation. Sydney: Macquarie University.

Ferrara, L., and Halvorsen, R. P. (2017). Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture* 16, 371–395. doi: 10.1075/gest.00001.fer

Ferrara, L., and Hodge, G. (2018). Language as description, indication, and depiction. *Front. Psychol.* 9:716. doi: 10.3389/fpsyg.2018.00716

Ferrara, L., and Johnston, T. (2014). Elaborating who's what: a study of constructed action and clause structure in Auslan (Australian Sign Language). *Austr. J. Ling.* 34, 193–215. doi: 10.1080/07268602.2014.887405

Fricke, E. (2008). *Grundlagen Einer Multimodalen Grammatik des Deutschen: Syntaktische Strukturen und Funktionen (Habilitation treatise)*. Frankfurt: European University Viadrina.

Fricke, E. (2012). *Grammatik Multimodal: Wie Wörter und Gesten Zusammenwirken*. Berlin: De Gruyter Mouton.

Fricke, E. (2013). "Towards a unified grammar of gesture and speech: a multimodal approach," in *Body—Language—Communication*, Vol. 1, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 733–754.

Gärdenfors, P. (2017). Demonstration and pantomime in the evolution of teaching. *Front. Psychol.* 8:415. doi: 10.3389/fpsyg.2017.00415

Golato, A. (2000). An innovative German quotative for reporting on embodied actions: Und ich so/und er so 'and i'm like/and he's like.'. *J. Pragmat.* 32, 29–54. doi: 10.1016/S0378-2166(99)00030-2

Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help us Think*. Cambridge, MA: Harvard University Press.

Goodman, N. (1968). *Languages of Art: An Approach to the Theory of Symbols*. Indianapolis: Bobbs-Merrill.

Goodwin, C. (2003). "Pointing as situated practice," in *Pointing: Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Psychology Press), 217–241.

Goodwin, M. H., and Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62, 51–76.

Green, J. (2014). *Drawn from the Ground: Sound, Sign and Inscription in Central Australian Sand Stories*. New York, NY: Cambridge University Press.

Green, J., and Wilkins, D. P. (2014). With or without speech: arandic sign language from Central Australia. *Austr. J. Ling.* 34, 234–261. doi: 10.1080/07268602.2014.887407

Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund: Lund University Press.

Hadar, U., and Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica* 115, 147–172. doi: 10.1515/semi.1997.115.1-2.147

Haiman, J. (1983). Iconic and economic motivation. *Language* 59, 781–819. doi: 10.2307/413373

Harrison, S. (2018). *The Impulse to Gesture: Where Language, Minds, and Bodies Intersect*. Cambridge: Cambridge University Press.

Hassemer, J. (2016). *Towards a Theory of Gesture Form Analysis: Imaginary Forms as Part of Gesture Conceptualisation, With Empirical Support from Motion-Capture Data*. Ph.D. dissertation. Aachen: RWTH Aachen University.

Hinnell, J. (2018). The multimodal marking of aspect: the case of five periphrastic auxiliary constructions in North American English. *Cogn. Ling.* 29, 773–806. doi: 10.1515/cog-2017-0009

Hinnell, J. (2019). The verbal-kinesic enactment of contrast in North American English. *Am. J. Semiot.* 35, 55–92. doi: 10.5840/ajs20198754

Hodge, G. (2014). *Patterns From a Signed Language Corpus: Clause-like Units in Auslan (Australian sign language)*. Ph.D. dissertation. Sydney: Macquarie University.

Hodge, G., and Cormier, K. (2019). Reported speech as enactment. *Ling. Typol.* 23, 185–196. doi: 10.1515/lingty-2019-0008

Hodge, G., and Ferrara, L. (2014). "Showing the story: enactment as performance in auslan narratives," in *Selected Papers from the 44th Conference of the Australian Linguistic Society, 2013*, eds L. Gawne and J. Vaughan (Melbourne: University of Melbourne), 372–397.

Hodge, G., Ferrara, L., and Anible, B. D. (2019). The semiotic diversity of doing reference in a deaf signed language. *J. Pragmat.* 143, 33–53. doi: 10.1016/j.pragma.2019.01.025

Hodge, G., and Johnston, T. (2014). Points, depictions, gestures and enactment: partly lexical and non-lexical signs as core elements of single clause-like units in Auslan (Australian Sign Language). *Austr. J. Ling.* 34, 262–291. doi: 10.1080/07268602.2014.887408

Hoffmann, T. (2017). Multimodal constructs – multimodal constructions? The role of constructions in the working memory. *Ling. Vanguard* 3(s1):20160042. doi: 10.1515/lingvan-2016-0042

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495

Hsu, H.-C. (2019). "Speech-embedded non-verbal depictions: embeddedness and structural complexities," *Paper Presented at the 15th International Cognitive Linguistics Conference*, Vol. 2019, Nishinomiya.

Hsu, H.-C., Brône, G., and Feyaerts, K. (to appear). In other gestures: Multimodal iteration in cello master classes. *Ling. Vanguard*

Jakobson, R. (1966). "Quest for the essence of language," in *Roman Jakobson: On Language*, eds L. R. Waugh and M. Monville-Burston (Cambridge, MA: Harvard University Press), 407–421.

Janzen, T. (2017). Composite utterances in a signed language: topic constructions and perspective-taking in ASL. *Cogn. Ling.* 28, 511–538. doi: 10.1515/cog-2016-0121

Johnston, T. (1996). "Function and medium in the forms of linguistic expression found in a sign language," in *International Review of Sign Linguistics*, Vol. 1, eds W. H. Edmondson and R. B. Wilbur (Mahwah, NJ: Lawrence Erlbaum), 57–94.

Keevallik, L. (2010). Bodily quoting in dance correction. *Res. Lang. Soc. Inter.* 43, 401–426. doi: 10.1080/08351813.2010.518065

Keevallik, L. (2015). "Coordinating the temporalities of talk and dance," in *Temporality in Interaction*, eds A. Deppermann and S. Günthner (Amsterdam: John Benjamins), 309–336. doi: 10.1075/slsi.27.10kee

Keevallik, L. (2017). "Linking performances: the temporality of contrastive grammar," in *Linking Clauses and Actions in Social Interaction*, eds R. Laury, M. Etelämäki, and E. Couper-Kuhlen (Helsinki: Finnish Literature Society).

Keevallik, L. (2018). What does embodied interaction tell us about grammar? *Res. Lang. Soc. Inter.* 51, 1–21. doi: 10.1080/08351813.2018.1413887

Keevallik, L. (2020). "Multimodal noun phrases," in *The 'Noun Phrase' Across Languages: An Emergent Unit in Interaction*, eds T. Ono and S. A. Thompson (Amsterdam: John Benjamins), 154–177. doi: 10.1075/tsl.128.07kee

Kendon, A. (1972). "Some relationships between body motion and speech," in *Studies in Dyadic Communication*, eds A. W. Siegman and B. Pope (New York, NY: Pergamon Press), 177–210. doi: 10.1016/b978-0-08-015867-9.50013-7

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 (Berlin: De Gruyter Mouton), 207–227.

Kendon, A. (1988a). "How gestures can become like words," in *Cross-Cultural Perspectives in Non-VERBAL COMMUNICATION*, ed. F. Poyatos (Toronto: Hogrefe), 131–141.

Kendon, A. (1988b). *Sign Languages of Aboriginal Australia: Cultural, Semiotic and Communicative Perspectives*. Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35, 379–415.

Kita, S. (2000). "How representational gestures help speaking," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 162–185. doi: 10.1017/cbo9780511620850.011

Kita, S. (2003). "Pointing: a foundational building block of human communication," in *Pointing: Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Psychology Press), 1–8. doi: 10.1007/978-94-009-3593-8_1

Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059

Kok, K. I., and Cienki, A. (2016). Cognitive grammar and gesture: points of convergence, advances and challenges. *Cogn. Ling.* 27, 67–100. doi: 10.1515/cog-2015-0087

Ladewig, S. H. (2014). "Recurrent gestures," in *Body—Language—Communication*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1558–1574.

Ladewig, S. H. (2020). *Integrating Gestures: The Dimension of Multimodality in Cognitive Grammar*. Berlin: De Gruyter Mouton.

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Langacker, R. W. (2016). *Nominal Structure in Cognitive Grammar*. Lublin: Marie-Curie Skłodowska University Press.

Lehmann, C. (1988). "Towards a typology of clause linkage," in *Clause Combining in Grammar and Discourse*, eds J. Haiman and S. A. Thompson (Amsterdam: John Benjamins), 181–225. doi: 10.1075/tsl.18.09leh

Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

Macaulay, R. (2001). You're like 'why not?' The quotative expressions of Glasgow adolescents. *J. Socioling.* 5, 3–21. doi: 10.1111/1467-9481.00135

Mandel, M. (1977). "Iconic devices in American sign language," in *On the Other Hand: New Perspectives on American Sign Language*, ed. L. A. Friedman (New York, NY: Academic Press), 57–108.

McGregor, W. B. (1997). *Semiotic Grammar*. Oxford: Clarendon Press.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2013a). "The co-evolution of gesture and speech, and downstream consequences," in *Body—Language—Communication*, Vol. 1, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 480–512.

McNeill, D. (2013b). "The growth point hypothesis of language and gesture as a dynamic and integrated system," in *Body—Language—Communication*, Vol. 1, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 135–155.

McNeill, D., and Duncan, S. D. (2000). "Growth points in thinking for speaking," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 141–161. doi: 10.1017/cbo9780511620850.010

Miller, J., and Weinert, R. (1995). The function of LIKE in dialogue. *J. Pragmat.* 23, 365–393. doi: 10.1016/0378-2166(94)00044-F

Mittelberg, I. (2014). "Gestures and iconicity," in *Body—Language—Communication*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1712–1732.

Mittelberg, I., and Evola, V. (2014). "Iconic and representational gestures," in *Body—Language—Communication*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1732–1746.

Mittelberg, I., and Waugh, L. R. (2014). "Gestures and metonymy," in *Body—Language—Communication*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1747–1766.

Mondada, L. (2014). "Pointing, talk, and the bodies: reference and joint attention as embodied interactional achievements," in *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, eds M. Seyfeddinipur and M. Gullberg (Amsterdam: John Benjamins), 95–124. doi: 10.1075/z.188.06mon

Mondada, L. (2016). *Conventions for Multimodal Transcription*. Available online at: https://franz.unibas.ch/fileadmin/franz/user_upload/redaktion/Mondada_conv_multimodality.pdf (accessed December 3, 2016).

Mondada, L. (2019). Contemporary issues in conversation analysis: embodiment and materiality, multimodality and multisensoriality in social interaction. *J. Pragmat.* 145, 47–62. doi: 10.1016/j.pragma.2019.01.016

Müller, C. (1998a). "Iconicity and gesture," in *Oralité et Gestualité*, eds S. Santi, I. Guaïtella, C. Cave, and G. Konopczynski (Paris: L'Harmattan), 321–328.

Müller, C. (1998b). *Redebegleitende Gesten. Kulturgeschichte — Theorie — Sprachvergleich*. Berlin: Berlin Verlag Arno Spitz.

Müller, C. (2010). Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. *Sprache Liter.* 41, 37–68. doi: 10.1163/25890859-041-01-90000004

Müller, C. (2014). "Gestural modes of representation as techniques of depiction," in *Body—Language—Communication*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1687–1702.

Müller, C. (2018). Gesture and sign: cataclysmic break or dynamic relations? *Front. Psychol.* 9:1651. doi: 10.3389/fpsyg.2018.01651

Müller, C., Ladewig, S. H., and Bressem, J. (2013). "Gestures and speech from a linguistic perspective: a new field and its history," in *Body—Language—Communication*, Vol. 1, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 55–81.

Navarretta, C. (2015). "The functions of fillers, filled pauses and co-occurring gestures in Danish dyadic conversations," in *Proceedings from the 3rd European Symposium on Multimodal Communication*, Dublin.

Parrill, F. (2009). Dual viewpoint gestures. *Gesture* 9, 271–289. doi: 10.1075/gest.9.3.01par

Peirce, C. S. (1932). "The icon, index, and symbol," in *Collected Papers of Charles Sanders Peirce*, Vol. 2, eds C. Hartshorne and P. Weiss (Cambridge, MA: Harvard University Press), 156–173.

Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227

Ruth-Hirrel, L., and Wilcox, S. (2018). Speech-gesture constructions in cognitive grammar: the case of beats and points. *Cogn. Ling.* 29, 453–493. doi: 10.1515/cog-2017-0116

Sambre, P., Brône, G., & Wermuth, C. (2019). "Verbal genericity and null instantiation in Italian and German cut and break sequences: a multimodal socio-cognitive approach," *Paper Presented at the 15th International Cognitive Linguistics Conference*, Nishinomiya.

Schoonjans, S. (2017). Multimodal construction grammar issues are construction grammar issues. *Ling. Vanguard* 3(s1):20160050. doi: 10.1515/lingvan-2016-0050

Schoonjans, S., Brône, G., and Feyaerts, K. (2015). "Multimodalität in der Konstruktionsgrammatik: eine kritische Betrachtung illustriert anhand einer Gestikanalyse der Partikel einfach," in *Konstruktionsgrammatik V: Konstruktionen im Spannungsfeld von sequenziellen Mustern, kommunikativen Gattungen und Textsorten*, eds J. Bücker, S. Günthner, and W. Imo (Tübingen: Stauffenburg), 291–308.

Sidnell, J. (2006). Coordinating gesture, talk, and gaze in reenactments. *Res. Lang. Soc. Inter.* 39, 377–409. doi: 10.1207/s15327973rlsi3904_2

Simone, R. (ed.) (1995). *Iconicity in Language*. Amsterdam: John Benjamins.

Slama-Cazacu, T. (1976). "Nonverbal components in message sequence: "Mixed syntax."," in *Language and Man: Anthropological Issues*, eds W. C. McCormack and S. A. Wurm (Berlin: De Gruyter Mouton), 217–227. doi: 10.1515/9783112321454-015

Sowa, T. (2006). *Understanding Coverbal Iconic Gestures in Shape Descriptions*. Berlin: Akademische Verlagsgesellschaft.

Spronck, S., and Nikitina, T. (2019). Reported speech forms a dedicated syntactic domain. *Ling. Typol.* 23, 119–159. doi: 10.1515/lingty-2019-0005

Stec, K., Huiskes, M., and Redeker, G. (2015). Multimodal analysis of quotation in oral narratives. *Open Ling.* 1, 531–554. doi: 10.1515/opli-2015-0018

Stec, K., Huiskes, M., and Redeker, G. (2016). Multimodal quotation: role shift practices in spoken narratives. *J. Pragmat.* 104, 1–17. doi: 10.1016/j.pragma.2016.07.008

Steen, F., and Turner, M. (2013). "Multimodal construction grammar," in *Language and the Creative Mind*, eds M. Borkent, B. Dancygier, and J. Hinnell (Stanford, CA: CSLI Publications), 255–274.

Stokoe, W. C. (1960). *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf*. Buffalo, NY: University of Buffalo.

Stokoe, W. C. (2001). *Language in Hand: Why Sign Came Before Speech*. Washington, DC: Gallaudet University Press.

Streeck, J. (2002). Grammars, words, and embodied meanings: on the uses and evolution of so and like. *J. Commun.* 52, 581–596. doi: 10.1111/j.1460-2466.2002.tb02563.x

Streeck, J. (2008). Depicting by gestures. *Gesture* 8, 285–301.

Streeck, J. (2009). *Gesturecraft: The Manu-facture of Meaning*. Amsterdam: John Benjamins.

Sweetser, E., and Stec, K. (2016). "Maintaining multiple viewpoints with gaze," in *Viewpoint and the Fabric of Meaning*, eds B. Dancygier, W. Lu, and A. Verhagen (Berlin: De Gruyter Mouton), 237–258.

Tabacaru, S. (2014). *Humorous Implications and Meanings: A Multi-Modal Approach to Sarcasm in Interactional Humor*. Ph.D. dissertation. Lille: Université Charles de Gaulle — Lille III.

Tagliamonte, S., and Hudson, R. (1999). Be like et al. beyond America: the quotative system in British and Canadian youth. *J. Socioling.* 3, 147–172. doi: 10.1111/1467-9481.00070

Tannen, D. (2007). *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*, 2nd Edn. Cambridge: Cambridge University Press.

Taylor, M. (2007). *British Pantomime Performance*. Bristol: Intellect Books.

Tomasello, M. (2008). *The Origins of Human Communication*. Cambridge, MA: Massachusetts Institute of Technology Press.

Turner, M. (2017). Multimodal form-meaning pairs for blended classic joint attention. *Ling. Vanguard* 3(s1):20160043. doi: 10.1515/lingvan-2016-0043

Vandelanotte, L. (2009). *Speech and Thought Representation in English: A Cognitive-Functional Approach*. Berlin: De Gruyter Mouton.

Vandelanotte, L., and Davidse, K. (2009). The emergence and structure of be like and related quotatives: a constructional account. *Cogn. Ling.* 20, 777–807. doi: 10.1515/COGL.2009.032

Vermeerbergen, M. (2006). Past and current trends in sign language research. *Lang. Commun.* 26, 168–192. doi: 10.1016/j.langcom.2005.10.004

Wade, E., and Clark, H. H. (1993). Reproduction and demonstration in quotations. *J. Mem. Lang.* 32, 805–819. doi: 10.1006/jmla.1993.1040

Wilcox, S. (2004). Cognitive iconicity: conceptual spaces, meaning, and gesture in signed languages. *Cogn. Ling.* 15, 119–147.

Wilcox, S., and Occhino, C. (2016). Constructing signs: place as a symbolic structure in signed languages. *Cogn. Ling.* 27 doi: 10.1515/cog-2016-0003

Winter, B., Perlman, M., and Matlock, T. (2013). Using space to talk and gesture about numbers: evidence from the TV News Archive. *Gesture* 13, 377–408. doi: 10.1075/gest.13.3.06win

Ziem, A. (2017). Do we really need a multimodal construction grammar?. *Ling. Vanguard* 3(s1):20160095. doi: 10.1515/lingvan-2016-0095

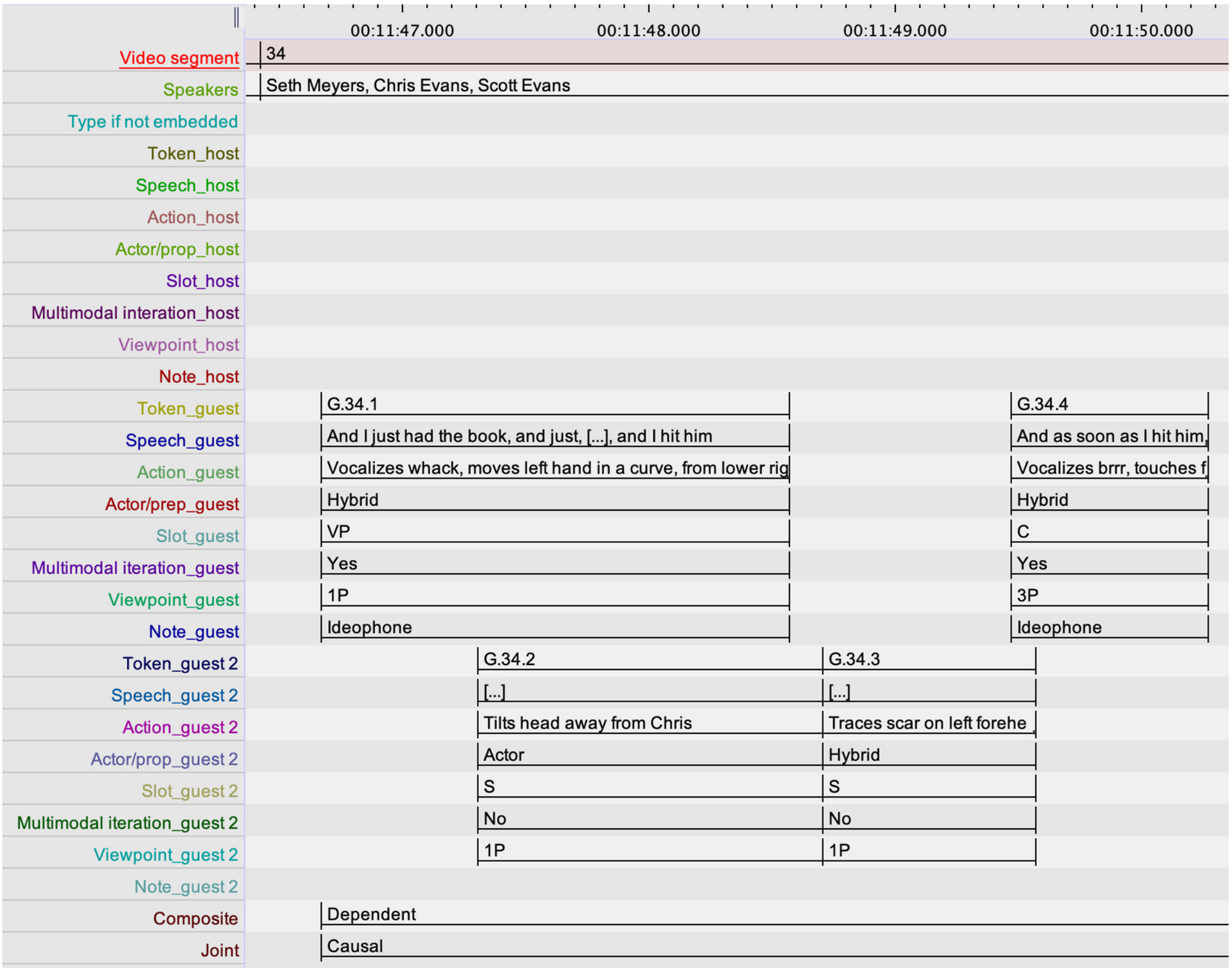
Zima, E. (2017). On the multimodality of [all the way from X PREP Y]. *Ling. Vanguard* 3(s1):20160055. doi: 10.1515/lingvan-2016-0055

Zima, E., and Bergs, A. (2017). Multimodality and construction grammar. *Ling. Vanguard* 3(s1):20161006. doi: 10.1515/lingvan-2016-1006

Zlatev, J. (2005). "What's in a schema? Bodily mimesis and the grounding of language, in *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. B. Hampe ed. Berlin: De Gruyter Mouton. 313–342. doi: 10.1515/9783110197532.4.313

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.

APPENDIX



00:11:47.000 00:11:48.000 00:11:49.000 00:11:50.000

Video segment	34			
Speakers	Seth Meyers, Chris Evans, Scott Evans			
Type if not embedded				
Token_host				
Speech_host				
Action_host				
Actor/prop_host				
Slot_host				
Multimodal interation_host				
Viewpoint_host				
Note_host				
Token_guest	G.34.1			G.34.4
Speech_guest	And I just had the book, and just, [...], and I hit him			And as soon as I hit him,
Action_guest	Vocalizes whack, moves left hand in a curve, from lower rig			Vocalizes brrr, touches f
Actor/prep_guest	Hybrid			Hybrid
Slot_guest	VP			C
Multimodal iteration_guest	Yes			Yes
Viewpoint_guest	1P			3P
Note_guest	Ideophone			Ideophone
Token_guest 2		G.34.2	G.34.3	
Speech_guest 2		[...]	[...]	
Action_guest 2		Tilts head away from Chris	Traces scar on left forehe	
Actor/prop_guest 2		Actor	Hybrid	
Slot_guest 2		S	S	
Multimodal iteration_guest 2		No	No	
Viewpoint_guest 2		1P	1P	
Note_guest 2				
Composite	Dependent			
Joint	Causal			

APPENDIX | Screenshot of annotation in ELAN (of Example 27).

REVIEW
published: 09 April 2021
doi: 10.3389/fpsyg.2021.650286

Integrating Embodied Cognition and Information Processing: A Combined Model of the Role of Gesture in Children's Mathematical Environments

Raychel Gordon and Geetha B. Ramani*

Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, United States

Children learn and use various strategies to solve math problems. One way children's math learning can be supported is through their use of and exposure to hand gestures. Children's self-produced gestures can reveal unique, math-relevant knowledge that is not contained in their speech. Additionally, these gestures can assist with their math learning and problem solving by supporting their cognitive processes, such as executive function. The gestures that children observe during math instructions are also linked to supporting cognition. Specifically, children are better able to learn, retain, and generalize knowledge about math when that information is presented within the gestures that accompany an instructor's speech. To date, no conceptual model provides an outline regarding how these gestures and the math environment are connected, nor how they may interact with children's underlying cognitive capacities such as their executive function. In this review, we propose a new model based on an integration of the information processing approach and theory of embodied cognition. We provide an in-depth review of the related literature and consider how prior research aligns with each link within the proposed model. Finally, we discuss the utility of the proposed model as it pertains to future research endeavors.

Keywords: gesture, math, executive function, children, learning

Edited by:
Naomi Sweller,
Macquarie University, Australia

Reviewed by:
Eliza Congdon,
Williams College, United States
Ruth Church,
Northeastern Illinois University,
United States

****Correspondence:***
Raychel Gordon
rjgordon@umd.edu

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

Received: *06 January 2021*
Accepted: *08 March 2021*
Published: *09 April 2021*

Citation:
Gordon R and Ramani GB (2021) Integrating Embodied Cognition and Information Processing: A Combined Model of the Role of Gesture in Children's Mathematical Environments . Front. Psychol. 12:650286. doi: 10.3389/fpsyg.2021.650286

INTRODUCTION

Hand gestures are used in a variety of mathematical contexts by children and adults alike. These gestures include both directed, meaningful movements intended to convey information, as well as less formal shifting of the hands that occurs naturally alongside conversation. Children use gestures to represent information, enhance conversation, and even support their own cognition (for a review, see Goldin-Meadow, 2009). Children's self-produced gestures (e.g., Broaders et al., 2007; Cook et al., 2008) as well as the gestures they see teachers use during instruction (e.g., Singer and Goldin-Meadow, 2005) have been shown to support their math learning. Given that children's early math knowledge has been consistently linked with their later math achievement (Claesens and Engel, 2013; Geary et al., 2013; Watts et al., 2014; Geary and Vanmarle, 2016), factors related to children's math understanding, like gesture, are important to understand.

One way that gestures can support mathematical learning is through their ability to aid children's cognition (Goldin-Meadow et al., 2001; Cook et al., 2012). Specifically, gesture can be linked to different components of domain-general abilities, such as executive function (EF). EF refers to the cognitive control processes that coordinate sub-processes such as attention shifting, working memory, and inhibitory control (e.g., Bull and Lee, 2014). Given the association between children's EF and math abilities (see Clements et al., 2016 for a review), as well as gestures supporting math through EF, we propose that it is important to study these factors together.

In this review, we discuss how children's use of and exposure to hand gestures during math contexts can shape their learning. Within the first section, we provide background information on two prevalent, but separate theories; information processing approach and theory of embodied cognition. We assert that combining these theories allows for a better understanding of the dynamic relations of gesture, EF, and children's mathematical learning. Thus, we propose a new model based on a combination of central tenets from these two theories. Next, we review relevant literature, and discuss how these results may be interpreted within the proposed model. In the final sections, we present opportunities for future empirical work. This paper serves as a unique examination of prior research through the lens of a unified model.

Overview of Gestures and Gesture Theories

Gestures are dynamic hand and body movements which accompany language. They can occur spontaneously or intentionally, and oftentimes provide different yet complementary information to a person's speech (Church and Goldin-Meadow, 1986; McNeill, 1992; Church, 1999). A speaker's gestures can facilitate listener's comprehension (for a meta-analysis, see Hostetter, 2011) and improve overall communication compared to speech alone (Church et al., 2000). Therefore, information provided by speakers' gestures are useful to those who see them.

Self-produced gestures can serve an important, internal purpose for the user as well. Hand gestures allow a speaker to simultaneously process their thoughts and put them into communicative form (McNeill, 1992; Krauss, 1998). People continue to gesture even when no one is watching (Krauss et al., 1995; Alibali et al., 2001). Research has shown that congenitally blind speakers use gestures even when they are communicating with a blind listener (Iverson and Goldin-Meadow, 1998), suggesting even those who have never seen gestures modeled in communication will use them too. Thus, gestures appear to support internal mechanisms of communication and cognition.

Due to the prevalence of gestures across ages, contexts, and domains, numerous theoretical models have been created to account for their communicative and cognitive functions. Each theory understandably overlaps in part with another; however, each one also provides complementary information explaining new contexts, factors, and functions. For example, frameworks that focus on what we can uncover about the speaker (Goldin-Meadow, 2003), or where these gestures emerge from (Gesture as Simulated Action framework, GSA; Hostetter and Alibali, 2008) both provide insight into how gestures relate to and shape underlying cognitive processes. Furthermore, the GSA framework builds upon another foundational idea that these cognitive processes are rooted within the environment (Embodied cognition, Barsalou, 1999). Gestures have also previously been considered under theories of cognition, such as Cognitive Load Theory (Sweller, 1988). This framework provides an explanation that self-produced gestures reduce "cognitive load," a mechanism that is often considered as one of the main roles of gestures. Each of these, as well as other gesture-related frameworks, provide unique and compelling explanations of the distinctive roles of gestures as they relate to a particular set of circumstances. However, these models do not consider the specific role of gestures in mathematical contexts. Our goal was to create a model based on the growing literature regarding the benefits of gestures, both produced and seen by children, during math environments.

A New Model of Gesture for Math Learning

We propose that the literature is best supported by a model that integrates two previously established frameworks. First is the information processing approach, commonly used within math research to represent how information moves through each component of human cognition during problem solving and learning. Second is the theory of embodied cognition, the basis of many gesture theories. This framework provides our model's infrastructure, as it articulates the importance of human-cognition being situated within a body, further encompassed in an active, stimuli-ridden environment.

Information Processing Approach

One way to conceptualize how children solve math problems and learn math related content is the Information Processing Approach (IP; e.g. Pellegrino and Goldman, 1987). This is not a single theory, rather an umbrella term for approaches which explains human cognition as a system that processes stimuli input from the environment and delivers a variety of outputs. The IP model suggests that learning occurs via a flow of information through a series of memory stores and processes. These distinctive elements in the IP approach can be conceptualized as the subcomponents of EF (adapted from Lutz and Huitt, 2003). Input is received from stimuli in the environment by way of the sensory registry. Attention is directed to fixate on relevant information, which progresses to working memory, a short-term store where information is held and processed for use in further cognitive tasks (Gathercole, 1998). Working memory is responsible for determining what information is important, choosing and enacting problem-solving strategies, and coming to a solution. Ultimately, information will be either be forgotten or encoded and stored in long-term memory for retrieval at a later time.

Although the IP framework can be broadly applied to represent children's math problem solving and learning, there are ways in which it could be further specified. First, a framework that focuses on both visual and auditory math-specific input could help to better understand how this input is relevant for children's math abilities and learning. Second, when investigating the role of gesture for children's early mathematics, it is important for a framework to include the body itself. While the IP model describes the cognitive processes, it does not explain any co-occurring physical behaviors. Thus, this framework cannot adequately account for the gesture-specific benefits that may occur within a math-related context. The question remains open as to how to model the role the learner's body, and the different types of math stimuli (words and gestures) within the environment.

Embodied Cognition

One theory that provides insight into these two components is embodied cognition (EC). While EC has been conceptualized in various ways, each adaptation generally emphasizes the body and stimuli within the surrounding environment as important to cognition (e.g., Barsalou, 1999; Clark, 1999; Shapiro, 2019). Here, we outline Wilson's (2002) presentation of EC. Specifically, she conveys six central claims of EC, three of which outline the importance of considering cognition as a situated process, and the other three focus on the importance of the body as a tool for cognition.

The first claim stipulates that cognition is situated. In other words, cognitive processing occurs in parallel with the task-relevant inputs and outputs from the environment. Thus, cognition cannot be separated from interplay between perception of the environment and subsequent actions taken. The second claim is that cognition is "time pressured," where cognitive processing requires real-time responses to the stimuli in their environment. Lastly, Wilson's fourth claim states that the environment is an important part of the cognitive system. Though similar to the first claim, Wilson outlines that since the reception of stimuli, cognitive processes, and behavioral responses are cyclical in nature, each of these components cannot be considered alone.

Wilson's claims three, five, and six all focus on the role of the individual's body in cognition. Claim three emphasizes that humans tend to off-load cognitive work externally in strategic ways. Wilson provides finger counting as an example, indicating this gesture can be used as a representation of relevant numeric information (e.g., linking number words to objects to keep track of quantity). Thus, offloading is a critical cognitive function that helps the speaker reason and express thoughts. The fifth claim states that cognition's primary function is for action. Meaning, a person's perception of the world as well as their concepts and memory are both "for" and "formulated by" their behaviors. Lastly, claim six says that off-line cognition is "body-based." Wilson's conceptualization of off-line cognitive processes involves any that are separable from the time-sensitive environment. Importantly, though they are distinct from the environment itself, the processes within the mind are inevitably tied to cognitive mechanisms that were originally designed for external behaviors, such as sensory processing and motor control.

The critical takeaway from Wilson's presentation of EC is that both the body and environment are integral to cognition. Her representation of EC underscores how embodied practices can result in an offloading of cognitive load. Based on how EC provides the important contribution of the body and the environment, and a focus on how cognition may be offloaded, we propose a model combining central tenets from both IP and EC.

The Proposed Model

The proposed model contains aspects from EC and IP, and specific contextual elements of gestures within the mathematical environment (**Figure 1**).

The proposed model is unique in its applicability to different math domains. For example, during a lesson on addition, math input could include a teacher's speech and gestures in reference to an equation on the chalkboard, while the output could be children's verbal and gestural response and explanations. In another context where a younger child is counting a set of objects, their math-input could be instructions and countable objects, and their output may include them pointing and counting out loud. Thus, there are numerous opportunities for applying the model for research by specifying the components (learner, input, and output) within a math environment.

Notably, our model also does not differentiate between perceived speech and gestures. Instead, it includes a unified representation of math-input. We base this combined representation on research showing that simultaneous presentation of these two observed modalities can be beneficial for children (Congdon et al., 2017). However, children's math-output is differentiated in the model because the literature (reviewed in subsequent sections) suggests children's gesture and speech often contain different but complementary information. For example, recent work supports separation of math output by modality, given that temporal-synchrony of self-produced gestures and speech does not relate to learning and retention for children in the same way that observed gestures do (Wakefield et al., 2021).

Incorporating gesture as input and output separately allows the model to be adapted in two critical ways. First, it can be applied to different mathematical domains (e.g., cardinality, algebra, fractions, etc.), such that the input and output can vary by content. Second, the model can be used to understand a broad range of differences in children's general EF abilities and math knowledge specifically. This is of particular importance given children are found to be adaptive in their responses to math problems (Siegler et al., 1996), and that the strategies children display may differ between their speech and gesture (Goldin-Meadow et al., 1993b).

Consider the example of a child solving the problem 3+2; if they have the answer memorized, they may quickly answer "5!" using a direct retrieval strategy of relevant math knowledge. A second child, who has only learned about addition principles generally, would likely respond differently to the same problem. They may use a backup strategy (i.e., any method other than retrieval), such as holding up three fingers then extending two more while counting on "4…5! The answer is 5!." The proposed model highlights how these children's individual differences in math knowledge could impact their use of self-produced gestures, and would allow researchers to explore the theoretical

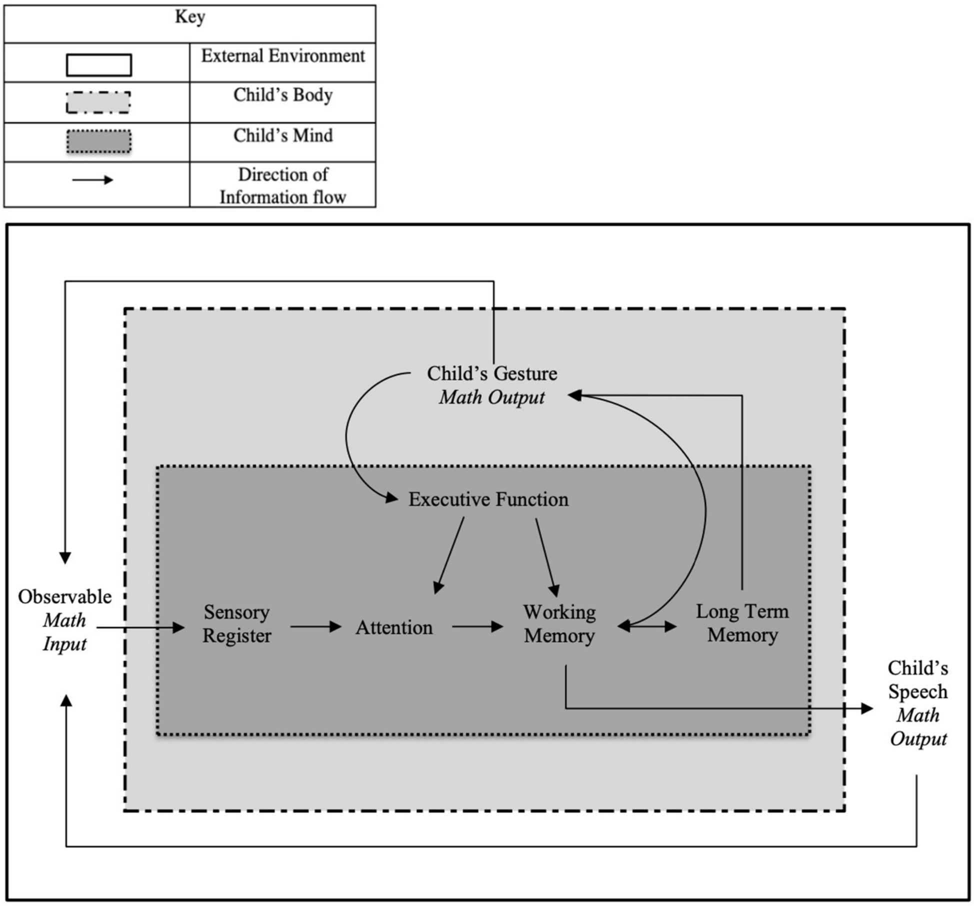


FIGURE 1 | Combined model of information processing and embodied cognition.

implications of how these strategies connect to their subsequent math abilities and later learning.

In addition to understanding the connection to math knowledge, an additional goal of the proposed model is to explain how gestures may be beneficial for EF and its subcomponents. EF includes three separate, but interrelated processes; attention shifting, inhibitory control, and working memory (Miyake et al., 2000). While EF is often discussed as a multidimensional construct, there is also evidence of unidimensionality in early childhood (Wiebe et al., 2008; Hughes et al., 2010). This makes it difficult to determine empirically whether the benefits of gesture for children relate to EF broadly, or to one specific sub-component. For example, it is common within the gesture literature to discuss gestures as providing a reduction in "cognitive load" (Goldin-Meadow et al., 2001) or linking them to executive function (EF) broadly (O'Neill and Miller, 2013). As such, connections between sub-components of EF and gesture are represented in the current review based on how they are discussed within their respective studies. The implications of this approach are reviewed in the discussion.

In sum, when studying the role of gesture in math environments, we propose a combined model of the IP approach and theory of EC. By establishing the pattern of information flow from the IP model within specific embodied locales and conventions of EC, this new model provides a dynamic representation of the cognitive impact of gestures in a mathematical environment. The connections between children's domain-specific knowledge (stored in long-term memory) to their self-produced gestures are illustrated within the model itself, as is an additional pathway between math-input and children's EF. Thus, both types of gestures are connected with children's cognition. Each of these connections will be considered through a review of the current literature.

Goals of the Current Review

There are two primary goals of the current review. First, we will review empirical work on the relations between children's mathematics ability and EF; mathematics ability and gesture use; as well as their gesture use and EF. Each of these will be discussed in terms of how this research fits within the proposed model. The second goal is to address any remaining gaps within the literature in order to demonstrate how the proposed model lays the foundation for future research to examine the mechanistic role that gesture may play in children's math learning.

Search Methodology

The present review is focused on connecting three separate, but related, bodies of literature: the gestures which children see and use, their EF abilities, and their mathematics knowledge. In the present review, less time is spent on connections between children's mathematics and EF abilities, given the numerous reviews of this topic that are readily available (for reviews, see Bull and Espy, 2006; Bull and Lee, 2014; and Cragg and Gilmore, 2014). Each of the subsequent components (gesture and math, gesture and EF) were investigated in a review conducted with APA PsycInfo database and Google Scholar in February through March of 2020. Follow-up searches were conducted in June-July of 2020. Relevant articles that came to our attention after our two initial searches were also included. In order to be included, studies must have been (1) published in English; (2) reports of original research, conceptualization of theory, or related reviews of literature published in peer-reviewed journals; (3) focused primarily on outcomes with children. Each search was conducted with separate keyword searches. Math and executive function were searched along with keyword combinations including but not limited to children, learning, and review. Math and gesture were searched along with keyword combinations including but not limited to children, learning, education, instruction, teacher, and review. Lastly, gesture and executive function were searched along with keyword combinations including but not limited to children, individual differences, working memory, attention, inhibition, childhood, and review.

MATH AND EXECUTIVE FUNCTION

Numerous studies have demonstrated a relation between children's mathematics ability and their EF (for reviews, see Bull and Espy, 2006; Bull and Lee, 2014; Cragg and Gilmore, 2014; Jacob and Parkinson, 2015; Peng et al., 2016). Broadly, this relation is consistent across different mathematical areas, including early numerical tasks, arithmetic problems, word problems, and standardized math measures (e.g., Lee et al., 2009, Bull et al., 2011; Van der Ven et al., 2012). It is critical to note that in both empirical and applied settings, EF has been conceptualized in numerous ways with researchers using a variety of assessment measures. As a result, empirical work on relations between math and EF are extensive and this literature has been previously reviewed as noted. Therefore, the focus of this section is to briefly summarize this research to demonstrate how representation of EF within the proposed model provides a specific operational system that is firmly connected to math contexts throughout childhood.

Cross-sectional correlational research has shown that different sub-components of EF are related to children's mathematical performance. For example, research indicates that working memory abilities are related to a range of mathematical tasks, such as early numeracy abilities (Kroesbergen et al., 2009), arithmetic achievement (Navarro et al., 2011), problem solving more broadly (Swanson, 2004), written and verbal calculation (Andersson, 2008), as well as mathematical word problem accuracy (Andersson, 2007; Zheng et al., 2011). Similar findings have shown connections between children's inhibitory control abilities and their math performance and achievement (Espy et al., 2004; Brock et al., 2009; Gilmore et al., 2013). There is additional evidence that inhibitory control, attention shifting, and working memory independently account for separate variance in children's math ability (Bull and Scerif, 2001). Further, when different sub-components of EF were examined in parallel, unique contributions of each on children's math ability were still prevalent (e.g., Bull and Scerif, 2001; St Clair-Thompson and Gathercole, 2006; Kroesbergen et al., 2009). Thus, evidence demonstrates relations between all three sub-components of EF and mathematics in children, lending support to including these factors within our model.

As children's mathematical knowledge develops, the impact of EF ability on their learning and performance differs. For children, it appears that working memory is of particular importance. Specifically, both children's symbolic (Caviola et al., 2012) and non-symbolic math abilities (Xenidou-Dervou et al., 2013) are positively related to their working memory. Importantly, children appear to rely more on their working memory than adults while solving math problems (Cragg et al., 2017). This may be due in part to how children's enactment of strategies is a more active and less efficient process and so their ability to enact a problem-solving strategy may be more of a direct result of their EF abilities compared to adults. Further, different EF abilities may allow an individual to enact different mathematical strategies (Imbo and Vandierendonck, 2007). For example, first grade children with higher working memory abilities were found to use more correct and sophisticated strategies on arithmetic problems compared to children with lower working memory capacity (Geary et al., 2012). These findings suggest that the relevance, contribution, and demand of working memory and broader EF abilities may shift depending on both the mathematical content and children's task knowledge, which can impact overall task performance. Thus, individual variation in EF abilities is a critical component to include in a model of children's math performance and learning, which is reflected in the centralized location of the proposed model.

Lastly, longitudinal studies have shown that children's EF is not only predictive of later mathematics performance (Alloway and Alloway, 2010; Monette et al., 2011), but also of their growth in mathematical abilities (Geary, 2011; Clark et al., 2013; LeFevre et al., 2013). For example, in a study following children from kindergarten to third grade, working memory related to children's early and later number competencies, which contributed to their math achievement (Krajewski and Schneider, 2009). However, training studies have shown mixed results. Some studies have found that EF training can improve children's numerical knowledge (Holmes et al., 2009; St Clair-Thompson et al., 2010; Holmes and Gathercole, 2014; Ramani et al., 2017, 2019). For example, training WM improved kindergarten children's counting skills, and WM games that included both numerical and non-numerical information improved children's counting and numerical comparison skills (Kroesbergen et al., 2014). However, others have found that providing children with EF training does not necessarily improve their mathematical knowledge (Jaeggi et al., 2012; Shipstead et al., 2012; Karbach et al., 2015). These findings suggest varying levels of efficacy in EF training on improvements in mathematics and provide the first window of opportunity for future research using the proposed model.

Overall, there is consistent cross-sectional and longitudinal evidence of relations between EF and mathematical achievement in children. These connections are found in a variety of mathematics domains, and the individual differences in EF abilities can influence children's mathematical performance. However, experimental evidence demonstrating that training EF can improve children's mathematical knowledge is less consistent, although numerous studies have shown promising results.

GESTURE AND MATH

In this section, we review literature on two types of gestures included in our model. First, we outline literature pertaining to gestures used by other people, such as a teacher or experimenter, to explain or teach math concepts; included in the model's "Math Input" section. Second, literature regarding children's self-produced gestures is reviewed; included in the model's "Math Output" section. Studies in these areas establish two critical functions (represented by connected arrows in **Figure 1**). One function highlights how children's self-produced gestures may convey math information (stored within their memory), which assists with their cognitive processing (EF). A second connection between children's gesture math-output connects back to their math input, which allows for the possibility that children's gestures elicit math information from their environment. Each of these are functions are reviewed and discussed.

Math Input: Observing Other People's Gestures

Individuals who observe a speaker's gestures during a mathematical context can extract useful information (Goldin-Meadow et al., 1992; Alibali et al., 1997; Kelly and Church, 1998). No training is required to gather this information, as children are readily able to attend to information found uniquely in gesture (Kelly and Church, 1997). Therefore, gestures that occur within math environments are straightforward in their presentation yet are critical to understand.

Experimental studies have shown that watching gestures can support learning and generalization of math concepts. For example, Graham (1999) had 2–4-year-old children (*n* = 85) watch a puppet point while counting objects. When asked about the puppet's performance, children spoke about the puppet's gestures suggesting that from a young age, children are explicitly able to recognize gesture strategies (pointing) in a math environment. Alibali and DiRusso (1999) used a similar paradigm with preschoolers (*n* = 20; M age = 4.67), where a subset of children was asked to count aloud while watching a puppet gesture to keep track of the objects. These children made fewer counting errors compared to children who had no gesture supports (either their own, or the puppets). These studies illustrate how young children can benefit from receiving gestures as part of their math input.

Research has also examined how gesture input could benefit other domains of math. Valenzeno et al. (2003) worked with 25 preschool-age children (M age = 4.5 years) who watched videos of teachers providing instruction on symmetry in a speech alone, or in gesture plus speech. Children who saw the gesture plus speech instructions had higher posttest scores for this math-concept, compared to children who received instruction in speech alone. Thus, children who received math-input with gesture showed greater improvement in math knowledge compared to their peers who received speech alone.

Additionally, children are also able to detect information that is uniquely communicated through gestural math-input. Specifically, Goldin-Meadow et al. (1999) asked a group of teachers to give children (*n* = 49, M age = 9.83 years) lessons on mathematical equivalence[1]. Teachers were not specifically told to gesture, though they did gesture spontaneously during instruction. These gestures contained relevant problem-solving strategies, such as a v-handshape to group two numbers together visually that should be summed, or gesticulating a flat palm under one side of a problem and then the other to indicate equality. These gestures either reinforced the information in the teacher's speech (gesture-speech match) or contained different, but complementary information (gesture-speech mismatch). Overall, children were more likely to reiterate their teacher's speech if it was accompanied by a gesture. Critically, children were also found to be able to recognize information that was solely presented within a teacher's gesture. This suggests that children both notice and process the mathematical information presented uniquely by gestures.

Children's ability to perceive information from gesture is further supported by evidence from a bilingual sample (Church et al., 2004). In this study, 51 Spanish-speaking first grade students (M age = 7.0 years) were assigned either to a Spanish-speaking classroom in school, or to an English-speaking classroom. Students watched a video of an English-speaking teacher providing instructions either with or without gestures. These gestures were gesture-speech mismatches, such that they contained unique but complementary information to speech. Overall, children in both classrooms benefited from the inclusion of gesture during instruction, and Spanish-speaking children's learning in particular increased from 0 to 50%. This suggests an additional benefit of including gestures as math-input. Specifically, gestures may be a more universally accessible representation of math information, as its manual format is not tied to a language.

Singer and Goldin-Meadow (2005) continued to build off this line of inquiry using the mathematical equivalence paradigm. Specifically, 3rd and 4th grade children (*n* = 160) were taught problem-solving strategies either with no gesture, gesture-speech matches, or gesture-speech mismatches. Children were

[1] An example of a problem involving the concept of math equivalence: in the problem $4 + 5 + 6 = \_ + 6$, children must recognize that the equals sign represents that one side should be equal to the other side, and that the problem requires them to solve for the blank within the equation.

more likely to learn when their teacher's math-input contained one problem-solving strategy in speech, while simultaneously presenting a different strategy in gesture. This finding extends previous work by suggesting that the addition of gesture to speech is unique in its ability to present two math concepts simultaneously (one in each modality), which in turn facilitates learning. Therefore, the inclusion of gestures as an accessible, beneficial form of math-input is cemented in the model.

It is additionally important to review research on how gesture input may impact children's math knowledge. Cook et al. (2013) asked 7–10-year-old children ($n = 184$) to watch a video where an instructor provided a lesson on math equivalence either with or without gestures. Children completed both an immediate and delayed posttest to test for general learning and transfer. Compared to children who received instruction in speech alone, children who received gestural math-input performed better on both the immediate and later posttests, including a transfer of knowledge to new problems. Thus, children appear to gain knowledge quicker and to generalize knowledge better when that information is provided with supporting gestures, as opposed to speech alone. These findings provide insight into how the inclusion of gestural math-input could impact children's own math-output, such as their response to a later math test.

Additional work expanded on these results with a computerized avatar (Cook et al., 2017). Sixty-five children (M age = 9.0) watched as a computer avatar provided instruction on mathematical equivalence, either with or without accompanying gestures. Children who saw the gesturing avatar learned more, solved problems quicker, and were more likely to generalize their knowledge. Thus, children benefited from the addition of gesture regardless of whether their instructor was human or a computer avatar. These results reveal how gestural math-input can be expanded to include technology-based instruction to advance children's learning and generalization of knowledge. This emphasizes the connections within the proposed model regarding math-input to children's overall math understanding.

Together, these findings suggest that children notice, and process mathematical information provided in instructor's gesture. These gestures are found to enhance children's learning and support broader understanding through concept transfer and generalization. This literature is consistent with the proposed model; children receive math input from their instructor's gestures and speech, which supports their problem solving and later learning in the form of math-output.

However, it is also critical to understand the mechanisms by which gestures provide these supports. One study assessed this issue by manipulating whether task-objects were in view, and thus referenceable, by their subjects (Ping and Goldin-Meadow, 2008). Specifically, kindergarten and first-grade students ($n =$ 61, 5–7 years old) participated in Piagetian conservation tasks where they were shown two objects (e.g., two glasses with equal liquid) and were asked if they were equal. One of the objects was manipulated, (e.g., poured into a shorter glass), such that children's understanding of conservation could be assessed when asked to explain if they were equal now. Children then received instruction on conservation, either in speech alone or gesture-plus-speech, as well as either with or without the objects present. On average, children were more likely to learn from instruction that contained gesture-plus-speech, even when the objects themselves were not present. In other words, gestural math-input was helpful beyond the scope of referencing specific, concrete objects within children's environment. Thus, the function of gesture as math-input goes beyond simple attention direction or grounding of speech in the physical environment and has broader implications for children's learning.

Overall, the literature suggests that the gestures which children observe as math-input can directly support their math learning, which reinforces these connections in the proposed model. Children are better able to learn, retain, and generalize new information about math when their instructor uses both gestures and speech, compared to speech alone. When children cannot access math-information in their teacher's speech, gestures become even more important. These benefits extend beyond a simple direction of attention, as gestures continue to be beneficial even when the relevant items are not present.

Math Output: Children's Self-Produced Gestures

In the following section, we review literature on the self-produced gestures children use in math contexts to scaffold their own knowledge and learning. These gestures occur spontaneously (e.g., Crowder and Newman, 1993) or as resulting from explicit instruction (e.g., Alibali and Goldin-Meadow, 1993). In the proposed model, children's own gestures are linked to supporting their ongoing cognitive functions, while also producing a form of math output. This output can then be observed by teachers to continue to inform the child's math environment (e.g. Gibson et al., 2019). Each of these functions of children's self-produced gestures are examined in turn.

Self-produced gestures have been shown to reduce cognitive load during math contexts. This benefit of gesture was examined by Goldin-Meadow et al. (2001), who asked participants to solve and explain age-appropriate math problem (e.g., math equivalence problems for children, harder problems for adults). They were also asked to remember a string of letters or words while providing the explanation for their solution. Gesture was manipulated directly, such that participants were given instruction regarding whether gestures were permitted, or if they should keep their hands on the table. Both adults and children were able to remember significantly more of their list when they used gestures during their math explanations. This finding supports the inclusion of children's self-produced gestures within the model. Furthermore, the authors suggest that the observed cognitive benefit may be due in part to gestures' utility in reducing memory demands, which may additionally link self-produced gestures to the memory processes in children's minds. Thus, this study is discussed briefly a second time in relation to working memory.

Another study investigated how self-produced gestures may further support children's performance on a math task. Specifically, Gordon et al. (2019) investigated how preschool children's own gestures may support their knowledge and performance on a cardinality task. Results indicated that children's cardinality knowledge was positively related to their spontaneous gesture use, even while controlling for age. This relation was not just driven by children who had mastered cardinality; indeed, the same positive relation between gesture and cardinality knowledge existed for the subsample of children who were still learning principles of cardinality. Children were also found to gesture the most during parts of the task that were most difficult for them, subjectively, based on their current cardinality knowledge. This emphasizes the connection in the model between children's own gestures, their math knowledge in long-term memory, facilitated by the problem-solving abilities within other components of EF.

Based on the advantages of self-produced gestures, additional work considers how providing explicit gesture instruction or encouragement to children to may impact their performance or learning in math environments. Broaders et al. (2007) examined this phenomenon in two studies with 3rd and 4th grade children who were asked to solve math equivalence problems. In the first study, children were asked to explain their solutions to these problems either using specifically with gesture, specifically without any gesture, or heard no mention of gesture. Children who were told to gesture conveyed different information in this modality (i.e., gesture-speech mismatch), such that their math-output contained new and relevant information. Therefore, instructing the use of gesture can lead children to express math knowledge with their hands that may not otherwise be communicated with their speech. The authors also sought to address whether children who received this instruction would be more receptive to learning by testing a new set of 3rd and 4th graders using a similar protocol for their second study. Results indicated that instructing children to use gesture not only taps into their implicit math-knowledge, but also makes them more likely to learn. Taken together, these results highlight how a combination of direct instruction (math-input) and the resulting self-produced gesture (math-output) could impact later math learning; the overall goal of the proposed model.

To further parse apart the benefits of instructed gestures, Goldin-Meadow et al. (2009) investigated whether specific types of gestures were more advantageous than others. Third and fourth graders completed math equivalence problems and were assigned to one of three training groups: no-gesture, correct-gesture, or partially correct gesture[2]. Overall, children learned more when any gesture was used, regardless of whether the information it contained was mathematically correct. However, children who received correct-gesture training solved more problems correctly compared to the partially correct gesture group. This suggests that gestures which contain specific, correct math information are superior to other gestural types. Furthermore, children were able to verbalize the grouping strategy used in gesture without any direct instruction, indicating that children learned a strategy from their own gestures. Taken together, these results indicate that while any gesture may benefit children, instructing specific gestures that align with math-concepts could allow children to extract and learn that information. This further supports the proposed model; children's self-produced gestures, while labeled as a form of "math-output," have connections to and from the knowledge storage and EF processes within their minds. Thus, by providing instruction to children to self-produce a specific type of gesture, they may be able to tap into and build on task-relevant knowledge.

New research involving fMRI methods builds on the mounting evidence that providing instruction to children to use gesture improves their mathematics ability. Wakefield et al. (2019) worked with 7–9-year-old children who had engaged in the same mathematical equivalence training outlined in previous research (Cook et al., 2008; Goldin-Meadow et al., 2009). Children solved a series of these problems, then received training to express an equalizer strategy in either speech alone or speech plus gesture. Only children who had gotten all problems wrong initially then successfully solved at least half the problems after training were included in the final sample ($n = 20$). A week later, this sub-sample of children completed a short training refresher before participating in an fMRI session where they solved new mathematical equivalence problems. Results showed differences in neural activation during problem solving by training condition, such that children in the speech and gesture condition had greater activation of the motor regions of their brains compared to speech-alone. This indicates that training math concepts through self-produced gestures may have lasting neural impacts, even though children were unable to use gesture during the fMRI reading itself. Thus, the neurological research is consistent in its support for the pathways generated by the behavioral research for the proposed model.

However, it is essential to address whether these benefits are unique to gesture, or if any movement or action could render the same benefits. For example, could children use a bodily strategy consisting solely of actions and have the same mathematical benefits? Novack et al. (2014) explored this idea with 3rd grade children using the math equivalence paradigm. Children were taught to use either a physical action on objects, a concrete gesture which mimicked that action, or an abstract gesture while solving the problem. While each of these strategies lead to more learning, only children who used gestures were able to generalize their knowledge to successfully complete later problems. Therefore, given that it is gesture rather physical action that best assists learning and knowledge transfer, the current model provides a unique vantage point to delve further into how gesture confers these benefits.

Building off this line of work, Congdon et al. (2018) investigated how individual differences in children's math knowledge influenced their learning from gesture or action strategies. First grade children's initial measurement knowledge was assessed, after which they received one of four trainings for a measurement task. Half of the conditions used a physical stick above a ruler aligned with zero, the other half shifted over to align with a different whole number. Conditions were further split by action or gesture-based trainings; Action-based

[2] Children in the no-gesture group only had the opportunity to verbalize the relevant equivalence strategy. Children in the correct gesture group learned to use their fingers to group the two, specific numbers that should be added to get to the correct answer. Children in the partially correct gesture condition still used a grouping gesture but with two numbers that would not sum to the correct answer.

trainings with physical manipulatives to show children how the ruler segments could be used to count, and gesture-based training using a "pinching" gesture to highlight the relevant segments of the ruler. Children who used simpler strategies incorrectly during the initial measurement assessment benefited from the action training, but not the gesture training. However, children who initially used a more complex, but incorrect, strategy at pretest learned from both the training with actions and gestures. This finding highlights the importance of recognizing how and when gestures could be applied, as well as how individual differences in children's own math knowledge may influence the benefits of gesture. In particular, encouraging the use of gesture may help a child who has reached the particular level of underlying math knowledge, yet hinder another less-advanced child at the same time. Thus, our model centralizes the importance of gesture while also highlighting the importance of not separating the utility of the tool from its intended user.

In educational settings, it is also important to understand how children's self-produced gestures can provide information to an observer, and how this observer could provide additional math-relevant input. In their seminal work, Church and Goldin-Meadow (1986) examined 5–8-year-old children's speech-gesture mismatches to investigate whether these movements indexed their transitional knowledge. In the first study ($n = 28$), children participated in a series of Piagetian conservation tasks where an experimenter made visual transformations of two equivalent objects. Throughout the task, children were asked if the objects had the same amount and to provide an explanation after the transformation. Children were categorized as a conserver (e.g., recognized the key concept of conservation), partial conserver, or a non-conserver based on their explanations. Children's speech and gesture use were coded during their explanation to determine if they were a match or a mismatch[3]. On average, children who had more mismatches showed more complex knowledge in their gestures than their speech. Based on this finding, the authors conducted a second study where half of the children received direct instruction on the concept of equivalence while the other half were given the opportunity to physically manipulate the objects. Children who had more speech-gesture mismatches in their explanations were more likely to learn new information after training and benefited from the opportunity to play and manipulate the objects afterwards. In contrast, those children with more matches than mismatches did not show any additional benefits from explicit training or more informal contact with the objects.

These findings were further expanded upon by Perry et al. (1988), who sought to explore how spontaneous self-produced gestures used in math contexts could index children's "readiness" to learn new information. In a series of studies, they asked 9–12-year-old children to solve problems and explain their solutions related to concepts of mathematical equivalence and Piagetian conservation. In general, children's speech and gestures were more likely to match during the conceptually easier mathematical task (conservation), but more likely to mismatch during the more difficult mathematical task (mathematical equivalence). Additionally, the amount and the type of mismatches produced by children provided an index of their "readiness" to learn. Specifically, the authors suggest that children's math-output (gesture and speech) provides insight into their math knowledge, as well as whether they may be able to receive new math-input from their environment. Indeed, children's gesture and speech mismatches have been linked to their zone of proximal development (Goldin-Meadow et al., 1993a). In other words, their gestures may be used by adults to specifically calibrate future math-input to a child's individual level of understanding.

To further understand how children's self-produced gestures mark their conceptual knowledge, Garber et al. (1998) assessed the speech gesture mismatches produced by 4th grade children in their explanations of mathematical equivalence problems. Children subsequently were asked to judge the acceptability of a variety of other commonly used problem-solving strategies, some of which were incorrect. Overall, children gave the highest rating to strategies which contained conceptual elements that they had only indicated in their gestures during their initial explanations of how to solve equivalence problems. Thus, these children not only expressed knowledge uniquely in their gestures, but this knowledge was accessible when presented to them later as additional mathematical input. Therefore, by watching the gestures that children produce as a type of mathematical output, it is possible to map out what math concepts they may already have some knowledge of implicitly. Taken together, these studies findings are consistent with the proposed model; that the gestures which children produce as a form of math output are linked to the knowledge stored within their long-term memory.

These markers of conceptual knowledge are found for other domains of math knowledge too. Specifically, Gunderson et al. (2015) studied 3–5-year-old children's mismatches in the context of cardinality, an early math concept which involves an extended learning process. Children who were still in the process of learning about this concept were more than twice as accurate in their gesture responses compared to their speech. Moreover, the gestures children produced were more accurate when the information in their gestures was a mismatch with their speech. Therefore, even young children who are in the process of learning a basic numerical concept provide unique information in their gestures that is not otherwise found in their speech. This finding supports that the current model may be extended to consider mathematics more broadly, as the patterns and information in gestural mismatches appear in the form of gestural math-output with younger children as well.

There is also evidence of this phenomenon in manual languages, such as American Sign Language (ASL). Goldin-Meadow et al. (2012) examined how the gestures produced by ASL-signing deaf children ($n = 40$) in the previously explained mathematical equivalence paradigm predicted whether they would benefit from explicit instruction on those problems. In general, the children who produced more gesture-sign mismatches were more likely to succeed after instruction than

[3] In the original work, when children's speech and gesture contained different information, it was labeled as "discordant", and if they contained the same information it was termed "concordant". However, these terms are used less commonly today, and we have replaced them to be consistent in our terminology across reference of this concept.

those who did not. This adds to the evidence by suggesting that mismatches occur even within the same modality, and strengthens the claim that it is critical to observe children's gestures as a form of math-output regardless of the modality of their language. Additionally, this finding highlights that the proposed integrated model may be extended for populations who use manual languages as well, though future research is required to further support each proposed connection.

In addition to studying whether the knowledge children express in gesture can be made available to them, it is also important to understand whether an external observer is able to recognize the utility of children's gestures. In other words, how does the literature support the connection within the model between children's self-produced gestures and the math-input they receive? One such study investigated this connection by recruiting a set of teachers ($n = 8$) to work with 3rd and 4th graders ($n =38$) on mathematical equivalence problems (Goldin-Meadow and Singer, 2003). Specifically, each child completed a pre-test of six problems, and explained their solutions to an experimenter. The teacher watched this pretest to gain insight on the child's knowledge, but was given no information or instruction regarding gestures. Each teacher then provided instruction on a set of problems before the child completed another, comparable posttest. Results showed that teachers were more likely to have variation in their instructions (e.g., give additional strategies) to children who had used more gesture-speech mismatches during their initial explanations. Therefore, children's own gestures (math-output) inadvertently shaped their own learning environment by evoking further explanation and support from the teacher (math-input). Not only does this happen spontaneously, but research shows that when adults are instructed to watch children's gestures, it can amplify the amount of information they were able to glean from children's gestures (Kelly et al., 2002). Even when the instruction was subtle, included different domains of knowledge, or different aged children, these results held. Thus, it is both possible to pick up on the information children possess implicitly by watching their gestures, and respond to these gestures in ways that may specifically scaffold the children's knowledge. These findings strengthen the connection within the integrated model between children's own math-output informing new math-input.

In sum, prior research provides evidence that self-produced gestures may benefit children's own learning and problem solving. These studies support the proposed, integrated model in several specific ways. First, they emphasize the modeled connection between math-input in children's environment and the subsequent impacts the input has on their math performance and learning. Second, literature which uniquely considers spontaneous or instructed self-produced gestures allows for additional insight to be added to the model, such as the how individual differences in children's knowledge made lead to differences in children's use of gestures, or differences in the benefits of gesture use itself. The same results are not reported with similar methods which employ physical action, which suggests that these mechanisms are unique to gesture. Additionally, prior research underscores the importance of centralizing the child within the model, given that a learner's own math knowledge and cognitive abilities can change the utility of gesture. Lastly, there is evidence suggesting that children's gestures are an indicator of their knowledge, and that this form of math-output that can be used as a tool by adults. This crucial collection of studies provides the connection within our model between children's gestures as math-output impacting the mathematical input they receive from others. Taken together, these studies highlight the necessity of a model where children's self-produced gestures in math environments can be studied further.

GESTURE AND EXECUTIVE FUNCTION (EF)

Given the multi-faceted role of gesture in children's math environments, it is critical to examine how the current literature supports the model's proposed connections between gestures and children's EF. Research outside the domain of mathematics has linked gesture specifically to EF from an early age (e.g., gesture, language, and EF; Kuhn et al., 2014). As previously discussed, individual's gestures may show information about implicit knowledge that is not found in their speech (Broaders et al., 2007; Pine et al., 2007). By shifting this information outside the mind and onto the hands, gesture is commonly proposed as a mechanism by which the user can "lighten their cognitive load" (Goldin-Meadow et al., 2001; Wagner et al., 2004). The idea of cognitive load is often presented as an offloading of related memory resources. While previous work has not drawn explicit connections to components of EF, more recent work has begun to delineate how gesture may be related to each subcomponents of EF. Thus, in this section, we review the literature regarding gesture, and their implied or direct connections made to the subcomponents of EF presented within the integrated model.

Working Memory

Working memory is a limited capacity sub-system of EF where information is temporarily held and processed during problem solving. On average, children use more gestures when faced with an explicit working memory demand (Delgado et al., 2011). The mechanistic connections between working memory and gesture are commonly discussed within the math and gesture literature.

For example, the aforementioned study by Goldin-Meadow et al. (2001) examined how children's memory could be impacted if they used gesture during some parts of the common math-equivalence task, but then were told to keep their hands still during other parts. Results indicated that participants performed better on the memory task when they were able to use gesture. This suggests the use of gesture allowed for a reduction of working memory load, compared when participants had to speak without gesturing. The authors suggest the use of gesture allowed for a reduction in working memory demands, allowing for a greater allocation of cognitive resources for the memory task, thereby improving performance. This same finding was found with adults. Using an updated, age-appropriate set of math problems to solve and explain as well as a harder set of memory items, adults were told they were allowed to use gesture only on some of their explanations. Similar to the children, the adults' performance was better when they were able to use gesture compared to when they only used speech, suggesting that both children and adults who use gesture while they speak would benefit in a reduction of working memory demands (Goldin-Meadow et al., 2001; Wagner et al., 2004). Thus, the current model reflects the direct connection between children's gestures and their working memory.

Ping and Goldin-Meadow (2010) further explored the mechanisms underlying how gestures benefit working memory. In this study, 2nd and 3rd grade children (M age 8.75 years) watched as an experimenter perform Piagetian conservation transformations. Children were asked to remember a list of words, then turned around to explain conservation to a new experimenter at another table. The new table was either empty or had the same conservation objects. This manipulation was critical as it allowed the researchers to test whether the cognitive benefits of gesture were based in its bodily capacity to link to a specific object or location (e.g., Ballard et al., 1997; Glenberg and Robertson, 1999). However, children who used gestures during their conservation explanations performed better on the memory task even when the items were absent and could not be directly indexed by gesture. Therefore, the working memory benefits of gesture are not tied to any specific object or spatial relation within the external environment.

More recent research with adults emphasizes the specific connection between gesture and working memory. For example, adults who are asked to use gesture may experience differential working memory benefits depending on their initial working memory abilities (Marstaller and Burianová, 2013). Additional studies have shown that people who have either lower visuospatial or verbal working memory capacity tend to produce more gestures on average (Chu et al., 2014; Gillespie et al., 2014; Pouw et al., 2016), and those who have higher than average visuospatial working memory abilities seem to be better equipped to detect information conveyed in gesture (Wu and Coulson, 2014a,b; Özer and Göksun, 2020). Thus, the connection between gesture and working memory are well-established.

The results of these studies are represented in the proposed model. Specifically, the proposed model reflects the bidirectional flow of information processing between children's own gestures and their working memory. This highlights the critical question of whether individual working memory abilities change how children receive gesture based math-input, as well as whether an individual's propensity to gesture could be impacted by their working memory abilities. In other words, would a child's initial working memory ability explain variability in their subsequent use of gesture within a math task?

Currently, there is not enough work available to answer this question. However, one recent study sought to address the related issue of whether the flow of information processing should vary based on a child's initial domain-general cognitive abilities. Specifically, recent research with preschoolers ($n = 81$) found that their spontaneous gestures and working memory were related to their performance on an age-appropriate math task (Gordon et al., 2021). However, children's gestures were not significantly related to their working memory after controlling for age. This work leaves room for future research to investigate this dynamic relation in further detail.

Attention

Additional work has informed the connection in our model between gestures and attention, another sub-component of EF. Research with infants indicates that they attend to pointing gestures before 6-months of age (Rohlfing et al., 2012). Shortly after 1 year, they begin to make their own attention-directing gestures to convey and request information from other people in their environment (Tomasello et al., 2007; Kovács et al., 2014), suggesting at least a basic understanding of the attentional function of gesture. Therefore, within the proposed model, children could be expected to both use and recognize the utility of gesture as a tool for attention.

However, the primary function of gesture is not only to drive attention. For example, one of the previously described studies exposed children to math gestures that contained task-relevant information, but also that directed their attention to irrelevant components of the math problem (Goldin-Meadow et al., 2009). Results showed children who saw these partially-correct gestures still learn more than children who received no gestures at all, suggesting that even though their attention may have been drawn to less relevant components, the gestures still helped. Nevertheless, attention has still been added as its own separate component within the proposed model, given that children in this study still learned the most when they received a gesture that contained both the task-relevant strategy information and directed their attention to the relevant parts of the problem. Therefore, we still believed it is important to include within our model that gestural math-input can direct children's attention towards relevant information within their environment.

Recent research lends additional support to retaining attention in some way within the proposed model. Specifically, Wakefield et al. (2018) investigated how gestural input could change children's visual attention during math instruction. Eight- to ten-year-old children ($n = 50$) participated in the math equivalence paradigm and watched videos of a teacher's instruction in speech alone or in speech and gesture. Children's eye movements were captured using eye-tracking technology, and their learning progress as well as concept transfer was assessed. Children who received both speech and gesture instruction spent time looking at both the problem and the gestures. Additionally, children who received instructions with both speech and gesture were more likely to follow along visually[4]. Following along was uniquely predictive of learning for those in the speech and gesture condition. Therefore, gesture as math input appears to moderate the impact of visual attention on learning and provides additional support for the inclusion of a connection between gesture input and attention within the proposed model.

The current model also ties children's self-produced gestures to their attention. There are limited empirical examples that directly test how children's own gestures drive their attention in ways that impact their math output and learning. However, Alibali and Kita (2010) assessed whether prohibiting children's

[4] For example, the instructor says "one side equal to the other side" while pointing the specific sides of the problem referenced in speech, and the child switches their gaze to the indicated components of the problem.

gestures would result in a shift of focus away from task-relevant information, which provides equal insight into this part of the model. In this study, researchers asked whether prohibiting 50 children (M age = 6 years, 5 months) from gesturing in the standard Piagetian conservation task would cause them to shift focus away from the perceptual-motor information which is commonly expressed in gesture. At first, all children were allowed to explain the conservation task with gesture, and then half the children were prohibited from gesturing for the second round of explanations by wearing a muff on their hands. On average, children were more likely to focus on information that was not perceptually present when they did not have access to gesture. When they were allowed to gesture, their focus shifted to the perceptually present information instead. Taken together, the results indicate that children's own gestures highlighted information within their own environment, and this information could be used in further cognitive processing related to children's later output. Therefore, while the main mechanism underlying gesture is not attention, it is still an essential component of EF that is tied to gesture. As such, the connection between gesture and attention within the proposed model are supported.

It is important to recognize that the proposed model does not include one connection built within the literature. Specifically, it has been suggested that individual speakers have a threshold for producing gestures, and that it may be possible for speakers to take advantage of this threshold (either directly or implicitly) to reap the cognitive benefits of gesture (Alibali and Nathan, 2012), suggesting a possible connection between attention sub-component of EF to gesture directly. The GSA framework provides a theoretical outlines how self-produced gestures are a consequence of a speaker's activation of own motor system involved in both planning and producing speech (Hostetter and Alibali, 2008). Based on a review of the empirical and theoretical supports, there is not enough support within the literature to draw a direct line from children's own math gestures to their own attention. As such, the proposed model only represents a flow of information routed by proxy of children's broader EF processes.

Inhibitory Control

Although inhibitory control is an important component of EF, it is currently not included in our proposed model. This is, in part, because less is known about how gesture may impact or be impacted by a children's inhibitory control. Here, we briefly review two studies outside of the scope of the mathematics to highlight the potential for future research.

First, O'Neill and Miller (2013) examined preschool children's gestures (M age = 47 months) during a Dimensional Change Card Sort task. Children ($n = 41$) were asked first to sort cards based on a given rule (e.g., sort cards by color), then midway switched to sorting the cards by a new rule (e.g., sort by shape). To sort successfully, children must inhibit the first rule to sort by the new rule. In general, children who gestured more had higher performance. Similar to math tasks, children who used specific task-relevant gestures had higher performance compared to children who used less relevant gestures. In particular, the majority of differences were noted after the rule shift, which is when children would have needed to inhibit the old rule to implement the new rule.

Additional work with preschoolers using the same card sort task assessed whether a direct, causal relation existed between preschoolers' gestures and their scores on another version of the Dimensional Change Card Sort task (Rhoads et al., 2018). Specifically, preschoolers received training to use gesture as a support during the task to retain the specific dimensions they were using to sort. On average, children who were instructed to gesture showed improved sorting accuracy, and these instructions appeared to be particularly beneficial for younger children. These results suggest that instructing children to use gesture may boost their overall EF performance, or even lead to specific improvements in their inhibitory control abilities.

While these results occur outside of the domain of mathematics, they suggest that children's gestures may help to keep new rules in mind, inhibit an old rule, or some combination of the two. While the proposed model provides a breakdown of EF, and the information that flows between the sub-systems of attention shifting and working memory, further research needs to be conducted to better understand how to incorporate the third component of EF, inhibitory control, into the model.

DISCUSSION

In the current paper, we narrow our focus from the function of gesture across learning contexts broadly (e.g. Goldin-Meadow and Wagner, 2005) and present a new model regarding the role of gesture in math environments. The processes involved in math learning are well-modeled by the Information Processing Approach, however this approach is not able to fully explain the underlying mechanisms of gesture. Thus, we include tenets of Wilson's (2002) presentation of EC by modeling the mind within the body, and by extension the surrounding environment. This allows for a consideration of gestures as a form of math-input from the environment, as well as a form of math-output from children's own bodies. After the model presentation, we review the relevant literature on each of the model components. First, we briefly summarize the literature between math and EF, providing additional motivation for operationalization of IP into the sub-components of EF. Next, we review literature pertaining to gesture both as a form of math-input and math-output. Lastly, we summarize studies pertaining to the cognitive benefits of gestures, and how these relate to the sub-components of EF. Here, we outline the strengths and weaknesses of the proposed model and make recommendations for future research.

One strength of the proposed model is its direct expression of the connections that have been made separately, or alluded to, in previous literature. For example, the integrated model ties findings from studies of EF and math to those of gestures in a math context. In doing so, this new model presents a more holistic representation of the connections between gesture, and EF, and math. Specifically, given that EF and math abilities have been robustly linked throughout childhood (e.g., Bull and Espy, 2006), new studies should account for whether differences in EF's subcomponents change the contribution of gesture.

A related strength of the proposed model is its direct attribution of benefits of gesture directly to the specific, and separate components of EF (e.g., working memory, attention, inhibitory control) when necessary. The benefits of gesture are commonly described in terms of promoting conceptual change or providing cognitive supports. For example, self-produced gestures are often said to reduce cognitive burden, "thereby freeing up effort that can be allocated to other tasks" (Goldin-Meadow, 1999, p. 427). This reduction of "cognitive burden" or influence on is still broadly used to represent the complicated, tangle of cognitive processes that are relevant to discussing how, when, and why gestures are beneficial (e.g., Novack and Goldin-Meadow, 2017). While there is evidence of unidimensionality across EF constructs in infancy and early childhood (Wiebe et al., 2008; Hughes et al., 2010), much of the literature emphasizes the number of distinct and separate components of EF in later childhood and adulthood (see Baggetta and Alexander, 2016 for a review). Thus, this model is the first of its kind to outline the connections between gestures, math, and the potential of developing, multidimensional components of EF for children.

Another strength of the model is its capacity to represent how individual differences may impact the role gesture. A recent meta-analysis investigating the role of observed and produced gestures in comprehension found that while gestures are generally beneficial to comprehension, they are most beneficial when a learner produces gesture themselves (Dargue et al., 2019). Indeed, there are even times where gestures do not promote learning (see Goldin-Meadow, 2010 for a review). Therefore, the proposed model is unique in that it highlights how variation at the core of the model (e.g., the learner's EF abilities, math-knowledge stored in their long-term memory, and other factors which shape these capacities) will change when and for whom gestures will promote learning.

In addition to these strengths, there are several areas for future research that this model helps to identify. Specifically, the proposed model is primarily informed by gesture instruction and gesture use during two mathematical concepts, the mathematical-equivalence paradigm (Perry et al., 1988) and Piagetian conservation tasks (Church and Goldin-Meadow, 1986). To date, many math-gesture researchers have chosen to use these paradigms as they have been shown to produce natural and relevant gestures. Therefore, our model is heavily informed by studies which have repeatedly tested their questions within the same specific mathematical domain. As such, our model is limited in its scope in terms of representing gesture in a broader array of math contexts, ages, and levels of cognition. Future studies may examine how this model reflects gesture in mathematics more broadly. For example, while some studies have been conducted on early mathematical skills (counting and cardinality), more research on the benefits of gestures for foundational math knowledge is of particular importance given that children's early abilities are strongly linked to their later math achievement (Claesens and Engel, 2013; Geary et al., 2013; Watts et al., 2014). As such, it is imperative to understand how and when to use gesture in the mathematics classroom to best maximize academic achievement.

An additional consideration for the current model is how well it aligns with other proposed frameworks of gesture. While our model allows for gesture-speech mismatches produced or witnessed by children in a math environment, we do not center our model around them. However, we do not believe that the decentralization of gesture-speech mismatches in the proposed model conflicts with prior literature. For example, literature considering self-produced speech-gesture mismatches find that they are indicative of student's readiness to learn new information (Church and Goldin-Meadow, 1986). On the other hand, watching a teacher's mismatches may actually drive student's learning, compared to those who receive matching or no gestures (Singer and Goldin-Meadow, 2005). Thus, we argue it is not just that speech-gesture mismatches contribute to conceptual change that is important. Instead, our model prominently features the distinct pathways by which these mismatches could impact children's cognition, and how this impact may vary depending on the source.

One gap in the current model is its ability to address the neural underpinnings of gesture (e.g., Wakefield et al., 2013, 2019). This line of work is imperative and may provide additional insight regarding how each related brain region could play a role in learning. However, the proposed model does not provide a basis for studies considering detailed neurophysiological components. This is not to say that the results of these studies could not be thought of in parallel with the behavioral measures outlined within the proposed model. Rather, it is our goal to provide an accurate representation both in terms of the model's primary objective, as well as its scope. As such, while support for the proposed model may be further strengthened by neuroscience methods, it is possible that a more precise neural model of gesture use may be needed.

The implications and opportunities for future research within this domain are broad. Specifically, there are several questions remaining to be answered: How do the individual differences in children's EF impact their use gestures during math tasks? Additionally, how does children's level of math knowledge impact their EF, self-produced gesture, or the interaction between the two? Do the types and rates of gesture vary as a function of problem difficulty, based on these individual differences? How does the nature of these relations change as children's math knowledge grows, and the specific content they are learning changes? Although each of these questions are motivated from the substantial research on children's gesture, mathematics, and executive function, key information is still missing.

As discussed previously, one gap in the literature is how children's inhibitory control may be linked to the mechanisms and benefits of gestures. Math contexts are particularly useful to study how children employ their inhibition abilities. For example, during problems solving children can inhibit old, ineffective, or incorrect strategies in lieu of new or correct strategies they have learned more recently (Siegler, 1996). Thus, future research could analyze how gesture may be used to support strategy inhibition during these critical learning periods. Additionally, children's spontaneous gestures could provide insight into their inhibition. If a child produces old, ineffective strategy knowledge in speech but newer strategy knowledge in gesture, this mismatch could imply that supporting their inhibitory control abilities would allow them to use the strategy knowledge they displayed in their gestures.

Additional research could be conducted to better support the proposed model's connection between the math knowledge stored within children's long-term memory and their self-produced gestures. The proposed model follows the current literature in that math knowledge can be displayed in children's self-produced gestures (e.g., Garber et al., 1998), and an assessment of this "implicit" knowledge can be used to determine whether children are ready to learn (Broaders et al., 2007), thereby leading to additional math-input. Thus, a unidirectional arrow leads from the information stored in children's long-term memory to their gestures, but these gestures loop out into the environment to inform their math-input. Thus, future research could directly investigate how this information changes depending on if children's gestures were spontaneous or the result of instruction. For example, while these types of gestures may appear to display similar information, it is possible that the underlying reason why gestures are generated in these circumstances could vary. Additionally, a child's propensity to gesture could differ based on the instructions they receive, and therefore the types and rates of self-produced gestures could also be expected to differ.

Relatedly, future research could examine the differences between when children receive specific instruction to use gestures themselves, compared to when they are just broadly exposed to gesture and mimic these movements independently. In other words, if a child is exposed to a particular type of gesture in a math context, what could we expect from them in later math settings? Would the presenter of that gesture matter in terms of whether it was a parent, teacher, or even a peer? In the event that children are told about the specific benefits of using gesture as a tool for math, would children use it in a way that helps them? Or would they over-employ gesture in ways that hinders performance? These are just a few of the many questions that the proposed model is uniquely suited to address. In particular, it allows future researchers to question how gesture-based math-input may facilitate learning, while simultaneously considering children's EF, math knowledge, and their own gestures and math-output.

CONCLUSION

The proposed model fuses central components of embodied cognition and information processing theories to highlight connections drawn in previous studies investigating gestures, EF, and math learning. Each component of this new model is outlined in a thorough review of the prior literature, through a combined lens of these two theories. Although there are several existing models of gestures and math learning, our model offers specific, novel avenues for future research. In particular, it provides a cohesive, theory driven representation of the role of gestures as they pertain to children's cognition within a math environment. In sum the proposed model provides future researchers with a theoretical foundation from which they may continue to understand the relations between gestures, EF, and children's math learning.

AUTHOR CONTRIBUTIONS

RG and GR developed the conceptual ideas. RG performed the literature search and drafted the manuscript. GR revised the manuscript and provided critical feedback for important intellectual content. All authors approved the submitted version.

FUNDING

This material is based upon work supported by the Faculty-Student Research Award (FSRA) from the Graduate School at the University of Maryland, College Park, and the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1840340.

REFERENCES

Alibali, M. W., and DiRusso, A. A. (1999). The function of gesture in learning to count: more than keeping track. *Cogn. Dev*. 14, 37–56. doi: 10.1016/S0885-2014(99)80017-3

Alibali, M. W., Flevares, L. M., and Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: do teachers have the upper hand?. *J. Educ. Psychol*. 89, 183–193. doi: 10.1037/0022-0663.89.1.183

Alibali, M. W., and Goldin-Meadow, S. (1993). Gesture–speech mismatch and mechanisms of learning: what the hands reveal about a child's state of mind. *Cognit. Psychol*. 25, 468–523 doi: 10.1006/cogp.1993.1012

Alibali, M. W., Heath, D. C., and Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *J. Mem. Lang*. 44, 169–188. doi: 10.1006/jmla.2000.2752

Alibali, M. W., and Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture* 10, 3–28. doi: 10.1075/gest.10.1.02ali

Alibali, M. W., and Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: evidence from learners' and teachers' gestures. *J. Learn. Sci*. 21, 247–286. doi: 10.1080/10508406.2011.611446

Alloway, T. P., and Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *J. Exp. Child Psychol.* 106, 20–29. doi: 10.1016/j.jecp.2009.11.003

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Appl. Cogn. Psychol.* 21, 1201–1216. doi: 10.1002/acp.1317

Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: The importance of central executive functions. *Br. J. Educ. Psychol.* 78, 181–203. doi: 10.1348/000709907X209854

Baggetta, P., and Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind Brain Educ.* 10, 10–33. doi: 10.1111/mbe.12100

Ballard, D. H., Hayhoe, M. M., Pook, P. K., and Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Brain Behav. Sci.* 20, 723–742. doi: 10.1017/S0140525X97001611

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci*. 22, 577–660. doi: 10.1017/S0140525X99002149

Broaders, S. C., Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *J. Exp. Psychol.* 136, 539–550. doi: 10.1037/0096-3445.136.4.539

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., and Grimm, K. J. (2009). The contributions of 'Hot' and 'cool' executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Child. Res. Q.* 24, 337–349. doi: 10.1016/j.ecresq.2009.06.001

Bull, R., and Espy, K. A. (2006). "Working memory, executive functioning, and children's mathematics," in *Working Memory and Education,* eds Pickering SJ (Burlington, MA: Academic Press), 94–123. doi: 10.1016/B978-012554465-8/50006-5

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., and Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: sources of variation in emergent mathematic achievement. *Dev. Sci.* 14, 679–692. doi: 10.1111/j.1467-7687.2010.01012.x

Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Perspect.* 8, 36–41. doi: 10.1111/cdep.12059

Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Dev. Neuropsychol.* 19, 273–293. doi: 10.1207/S15326942DN1903_3

Caviola, S., Mammarella, I. C., Cornoldi, C., and Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *J. Exp. Child Psychol.* 112, 141–160. doi: 10.1016/j.jecp.2012.02.005

Chu, M., Meyer, A., Foulkes, L., and Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *J. Exp. Psychol.* 143, 694–709. doi: 10.1037/a0033861

Church, R. B. (1999). Using gesture and speech to capture transitions in learning. *Cogn. Dev.* 14, 313–342. doi: 10.1016/S0885-2014(99)00007-6

Church, R. B., Ayman-Nolley, S., and Mahootian, S. (2004). The role of gesture in bilingual education: does gesture enhance learning?. *Int. J. Biling. Educ. Biling.* 7, 303–319. doi: 10.1080/13670050408667815

Church, R. B., and Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23, 43–71. doi: 10.1016/0010-0277(86)90053-3

Church, R. B., Kelly, S. D., and Lynch, K. (2000). Immediate memory for mismatched speech and representational gesture across development. *J. Nonverbal Behav.* 24, 151–174. doi: 10.1023/A:1006610013873

Claesens, A., and Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record.* 115, 1–29. Available online at: https://psycnet.apa.org/record/2013-18273-005

Clark, A. (1999). An embodied cognitive science?. *Trends Cogn. Sci.* 3, 345–351. doi: 10.1016/S1364-6613(99)01361-3

Clark, C. A., Sheffield, T. D., Wiebe, S. A., and Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Dev.* 84, 662–677. doi: 10.1111/j.1467-8624.2012.01854.x

Clements, D. H., Sarama, J., and Germeroth, C. (2016). Learning executive function and early mathematics: directions of causal relations. *Early Child. Res. Q.* 36, 79–90. doi: 10.1016/j.ecresq.2015.12.009

Congdon, E. L., Kwon, M. K., and Levine, S. C. (2018). Learning to measure through action and gesture: children's prior knowledge matters. *Cognition* 180, 182–190. doi: 10.1016/j.cognition.2018.07.002

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., and Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learn. Instr.* 50, 65–74. doi: 10.1016/j.learninstruc.2017.03.005

Cook, S. W., Duffy, R. G., and Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Dev.* 84, 1863–1871. doi: 10.1111/cdev.12097

Cook, S. W., Friedman, H. S., Duggan, K. A., Cui, J., and Popescu, V. (2017). Hand gesture and mathematics learning: lessons from an Avatar. *Cogn. Sci.* 41, 518–535. doi: 10.1111/cogs.12344

Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition* 106, 1047–1058. doi: 10.1016/j.cognition.2007.04.010

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Lang. Cogn. Process* 27, 594–610. doi: 10.1080/01690965.2011.567074

Cragg, L., and Gilmore, C. (2014). Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. *Trends Neurosci. Educ.* 3, 63–68. doi: 10.1016/j.tine.2013.12.001

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., and Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition* 162, 12–26. doi: 10.1016/j.cognition.2017.01.014

Crowder, E. M., and Newman, D. (1993). Telling what they know: the role of gesture and language in children's science explanations. *Pragmat. Cogn.* 1, 341–376. doi: 10.1075/pc.1.2.06cro

Dargue, N., Sweller, N., and Jones, M. P. (2019). When our hands help us understand: a meta-analysis into the effects of gesture on comprehension. *Psychol. Bull.* 145, 765–784. doi: 10.1037/bul0000202

Delgado, B., Gómez, J. C., and Sarriá, E. (2011). Pointing gestures as a cognitive tool in young children: experimental evidence. *J. Exp. Child Psychol.* 110, 299–312. doi: 10.1016/j.jecp.2011.04.010

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., and Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Dev. Neuropsychol.* 26, 465–486. doi: 10.1207/s15326942dn2601_6

Garber, P., Alibali, M. W., and Goldin-Meadow, S. (1998). Knowledge conveyed in gesture is not tied to the hands. *Child Dev.* 69, 75–84. doi: 10.2307/1132071

Gathercole, S. E. (1998). The development of memory. *J. Child Psychol. Psychiatry* 39, 3–27. doi: 10.1017/S0021963097001753

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510

Geary, D. C., Hoard, M. K., and Nugent, L. (2012). Independent contributions of the central executive, intelligence, and in-class attentive behavior to developmental change in the strategies used to solve addition problems. *J. Exp. Child Psychol.* 113, 49–65. doi: 10.1016/j.jecp.2012.03.003

Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE* 8:e54651. doi: 10.1371/journal.pone.0054651

Geary, D. C., and Vanmarle, K. (2016). Young children's core symbolic and nonsymbolic quantitative knowledge in the prediction of later mathematics achievement. *Dev. Psychol.* 52, 2130–2144. doi: 10.1037/dev0000214

Gibson, D. J., Gunderson, E. A., Spaepen, E., Levine, S. C., and Goldin-Meadow, S. (2019). Number gestures predict learning of number words. *Dev. Sci.* 22:e12791. doi: 10.1111/desc.12791

Gillespie, M., James, A. N., Federmeier, K. D., and Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: evidence from individual differences. *Cognition* 132, 174–180. doi: 10.1016/j.cognition.2014.03.012

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE* 8:e67374. doi: 10.1371/journal.pone.0067374

Glenberg, A. M., and Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Proc.* 28, 1–26. doi: 10.1080/01638539909545067

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends Cogn. Sci.* 3, 419-429.doi: 10.1016/S1364-6613(99)01397-2

Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think.* Cambridge, MA: Harvard University Press. doi: 10.1037/e413812005-377

Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child Dev. Perspect.* 3, 106–111. doi: 10.1111/j.1750-8606.2009.00088.x

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Lang. Cogn.* 2, 1–19. doi: 10.1515/langcog.2010.001

Goldin-Meadow, S., Alibali, M. W., and Church, R. B. (1993a). Transitions in concept acquisition: using the hand to read the mind. *Psychol. Rev.* 100, 279–297. doi: 10.1037/0033-295X.100.2.279

Goldin-Meadow, S., Cook, S. W., and Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychol. Sci.* 20, 267–272. doi: 10.1111/j.1467-9280.2009.02297.x

Goldin-Meadow, S., Kim, S., and Singer, M. (1999). What the teacher's hands tell the student's mind about math. *J. Educ. Psychol.* 91, 720–730. doi: 10.1037/0022-0663.91.4.720

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395

Goldin-Meadow, S., Nusbaum, H. C., Garber, P., and Church, R. B. (1993b). Transitions in learning: evidence for simultaneously activated strategies. *J. Exp. Psychol.* 19, 92–107. doi: 10.1037/0096-1523.19.1.92

Goldin-Meadow, S., Shield, A., Lenzen, D., Herzig, M., and Padden, C. (2012). The gestures ASL signers use tell us when they are ready to learn math. *Cognition* 123, 448–453. doi: 10.1016/j.cognition.2012.02.006

Goldin-Meadow, S., and Singer, M. A. (2003). From children's hands to adults' ears: gesture's role in the learning process. *Dev. Psychol*. 39, 509–520. doi: 10.1037/0012-1649.39.3.509

Goldin-Meadow, S., and Wagner, S. M. (2005). How our hands help us learn. *Trends Cogn. Sci.* 9, 234–241. doi: 10.1016/j.tics.2005.03.006

Goldin-Meadow, S., Wein, D., and Chang, C. (1992). Assessing knowledge through gesture: using children's hands to read their minds. *Cogn. Instr*. 9, 201–219. doi: 10.1207/s1532690xci0903_2

Gordon, R., Chernyak, N., and Cordes, S. (2019). Get to the point: preschoolers' spontaneous gesture use during a cardinality task. *Cogn. Dev*. 52:100818. doi: 10.1016/j.cogdev.2019.100818

Gordon, R., Scalise, N. R., and Ramani, G. B. (2021). Give yourself a hand: the role of gesture and working memory in preschoolers' numerical knowledge. *J. Exp. Child Psychol.* [Epub ahead of print].

Graham, T. A. (1999). The role of gesture in children's learning to count. *J. Exp. Child Psychol*. 74, 333–355. doi: 10.1006/jecp.1999.2520

Gunderson, E. A., Spaepen, E., Gibson, D., Goldin-Meadow, S., and Levine, S. C. (2015). Gesture as a window onto children's number knowledge. *Cognition* 144, 14–28. doi: 10.1016/j.cognition.2015.07.008

Holmes, J., and Gathercole, S. E. (2014). Taking working memory training from the laboratory into schools. *Educ. Psychol*. 34, 440–450. doi: 10.1080/01443410.2013.797338

Holmes, J., Gathercole, S. E., and Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Dev. Sci*. 12, F9–F15. doi: 10.1111/j.1467-7687.2009.00848.x

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol. Bull.* 137:297. doi: 10.1037/a0022128

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495

Hughes, C., Ensor, R., Wilson, A., and Graham, A. (2010). Tracking executive function across the transition to school: a latent variable approach. *Dev. Neuropsychol*. 35, 20–36. doi: 10.1080/87565640903325691

Imbo, I., and Vandierendonck, A. (2007). The development of strategy use in elementary school children: working memory and individual differences. *J. Exp. Child Psychol*. 96, 284–309. doi: 10.1016/j.jecp.2006.09.001

Iverson, J. M., and Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature* 396, 228–228. doi: 10.1038/24300

Jacob, R., and Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: a review. *Rev. Educ. Res*. 85, 512–552. doi: 10.3102/0034654314561338

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Shah, P. (2012). Cogmed and working memory training – current challenges and the search for underlying mechanisms. *J. Appl. Res. Mem. Cogn*. 1, 211–213. doi: 10.1016/j.jarmac.2012.07.002

Karbach, J., Strobach, T., and Schubert, T. (2015). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychol*. 21, 285–301. doi: 10.1080/09297049.2014.899336

Kelly, S. D., and Church, R. B. (1997). Can children detect conceptual information conveyed through other children's nonverbal behavior's. *Cogn. Instr*. 15, 107–134. doi: 10.1207/s1532690xci1501_4

Kelly, S. D., and Church, R. B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Dev*. 69, 85–93. doi: 10.1111/j.1467-8624.1998.tb06135.x

Kelly, S. D., Singer, M., Hicks, J., and Goldin-Meadow, S. (2002). A helping hand in assessing children's knowledge: instructing adults to attend to gesture. *Cogn. Instr*. 20, 1–26. doi: 10.1207/S1532690XCI2001_1

Kovács, A. M., Tauzin, T., Téglás, E., Gergely, G., and Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy,* 19, 543–557. doi: 10.1111/infa.12060

Krajewski, K., and Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: findings from a 3-year longitudinal study. *J. Exp. Child Psychol.* 103, 516–531. doi: 10.1016/j.jecp.2009.03.009

Krauss, R. M. (1998). Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* 7, 54–54. doi: 10.1111/1467-8721.ep13175642

Krauss, R. M., Dushay, R. A., Chen, Y., and Rauscher, F. (1995). The communicative value of conversational hand gesture. *J. Exp. Soc. Psychol*. 31, 533–552. doi: 10.1006/jesp.1995.1024

Kroesbergen, E. H., Van Luit, J. E. H., Van Lieshout, E. C. D. M., Van Loosbroek, E., and Van de Rijt, B. A. M. (2009). Individual differences in early numeracy: the role of executive functions and subitizing. *J. Psychoeduc. Assess*. 27, 226–236. doi: 10.1177/0734282908330586

Kroesbergen, E. H., van't Noordende, J. E., and Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychol.* 20, 23–37. doi: 10.1080/09297049.2012.736483

Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., Blair, C. B., and Family Life Project Key Investigators (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Dev*. 85, 1898–1914. doi: 10.1111/cdev.12249

Lee, K., Ng, E. L., and Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *J. Educ. Psychol*. 101, 373–387. doi: 10.1037/a0013843

LeFevre, J. A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S. L., et al. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *J. Exp. Child Psychol*. 114, 243–261. doi: 10.1016/j.jecp.2012.10.005

Lutz, S., and Huitt, W. (2003). "Information processing and memory: theory and applications," in *Educational Psychology Interactive* (Valdosta, GA: Valdosta State University), 1–17. Available online at: http://www.edpsycinteractive.org/papers/infoproc.pdf (accessed February, 2020).

Marstaller, L., and Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychon. Bull. Rev*. 20, 496–500. doi: 10.3758/s13423-012-0365-0

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol*. 41, 49–100. doi: 10.1006/cogp.1999.0734

Monette, S., Bigras, M., and Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *J. Exp. Child Psychol*. 109, 158–173. doi: 10.1016/j.jecp.2011.01.008

Navarro, J. I., Aguilar, M., Alcalde, C., Ruiz, G., Marchena, E., and Menacho, I. (2011). Inhibitory processes, working memory, phonological awareness, naming speed, and early arithmetic achievement. *Span. J. Psychol*. 14, 580–588. doi: 10.5209/rev_SJOP.2011.v14.n2.6

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., and Goldin-Meadow, S. (2014). From action to abstraction: using the hands to learn math. *Psychol. Sci*. 25, 903–910. doi: 10.1177/0956797613518351

Novack, M. A., and Goldin-Meadow, S. (2017). Gesture as representational action: a paper about function. *Psychon. Bull. Rev*. 24, 652–665. doi: 10.3758/s13423-016-1145-z

O'Neill, G., and Miller, P. H. (2013). A show of hands: relations between young children's gesturing and executive function. *Dev. Psychol*. 49, 1517–1528. doi: 10.1037/a0030241

Özer, D., and Göksun, T. (2020). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Lang. Cogn. Neurosci*. 35, 896–914. doi: 10.1080/23273798.2019.1703016

Pellegrino, J. W., and Goldman, S. R. (1987). Information processing and elementary mathematics. *J. Learn. Disabil.* 20, 23–32. doi: 10.1177/002221948702000105

Peng, P., Namkung, J., Barnes, M., and Sun, C. (2016). A meta-analysis of mathematics and working memory: moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *J. Educ. Psychol*. 108, 455–473. doi: 10.1037/edu0000079

Perry, M., Church, R., and Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cogn. Dev*. 3, 359–400. doi: 10.1016/0885-2014(88)90021-4

Pine, K. J., Lufkin, N., Kirk, E., and Messer, D. (2007). A microgenetic analysis of the relationship between speech and gesture in children: Evidence for semantic and temporal asynchrony. *Lang. Cogn. Process*. 22, 234–246. doi: 10.1080/01690960600630881

Ping, R., and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cogn. Sci*. 34, 602–619. doi: 10.1111/j.1551-6709.2010.01102.x

Ping, R. M., and Goldin-Meadow, S. (2008). Hands in the air: using ungrounded iconic gestures to teach children conservation of quantity. *Dev. Psychol*. 44, 1277–1287. doi: 10.1037/0012-1649.44.5.1277

Pouw, W. T., Mavilidi, M. F., Van Gog, T., and Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cogn. Process*. 17, 269–277. doi: 10.1007/s10339-016-0757-6

Ramani, G. B., Daubert, E. N., and Scalise, N. R. (2019). "Role of play and games in building children's foundational numerical knowledge," *in Cognitive Foundations for Improving Mathematical Learning*, eds D. C. Geary, D. B. Berch, and K. M. Koepke, (Cambridge, MA: Elsevier Academic Press), 69–90. doi: 10.1016/B978-0-12-815952-1.00003-7

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., and Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *J. Numerical Cogn.* 3, 468–495. doi: 10.5964/jnc.v3i2.31

Rhoads, C. L., Miller, P. H., and Jaeger, G. O. (2018). Put your hands up! Gesturing improves preschoolers' executive function. *J. Exp. Child Psychol*. 173, 41–58. doi: 10.1016/j.jecp.2018.03.010

Rohlfing, K. J., Longo, M. R., and Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Dev. Sci.* 15, 426–435. doi: 10.1111/j.1467-7687.2012.01139.x

Shapiro, L. (2019). *Embodied Cognition*. New York, NY: Routledge. doi: 10.4324/9781315180380

Shipstead, Z., Redick, T. S., and Engle, R. W. (2012). Is working memory training effective? *Psychol. Bull*. 138, 628–654. doi: 10.1037/a0027473

Siegler, R. S. (1996). *Emerging Minds: The Process of Change in Children's Thinking*. New York, NY: Oxford University Press. doi: 10.1093/oso/9780195077872.001.0001

Siegler, R. S., Adolph, K. E., and Lemaire, P. (1996). "Strategy choices across the life span," in *Implicit Memory and Metacognition,* ed Bossert O (New York, NY: Oxford University Press) 79–121.

Singer, M. A., and Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychol. Sci*. 16, 85–89. doi: 10.1111/j.0956-7976.2005.00786.x

St Clair-Thompson, H., Stevens, R., Hunt, A., and Bolder, E. (2010). Improving children's working memory and classroom performance. *Educ. Psychol*. 30, 203–219. doi: 10.1080/01443410903509259

St Clair-Thompson, H. L., and Gathercole, S. E. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *Quart. J. Exp. Psychol*. 59, 745–759. doi: 10.1080/17470210500162854

Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Mem. Cognit*. 32, 648–661. doi: 10.3758/BF03195856

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cogn. Sci*. 12, 257–285. doi: 10.1207/s15516709cog1202_4

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

Valenzeno, L., Alibali, M. W., and Klatzky, R. (2003). Teachers' gestures facilitate students' learning: a lesson in symmetry. *Contemp. Educ. Psychol*. 28, 187–204. doi: 10.1016/S0361-476X(02)00007-3

Van der Ven, S. H., Kroesbergen, E. H., Boom, J., and Leseman, P. P. (2012). The development of executive functions and early mathematics: a dynamic relationship. *Br. J. Educ. Psychol*. 82, 100–119. doi: 10.1111/j.2044-8279.2011.02035.x

Wagner, S. M., Nusbaum, H., and Goldin-Meadow, S. (2004). Probing the mental representation of gesture: is handwaving spatial? *J. Mem. Lang*. 50, 395–407. doi: 10.1016/j.jml.2004.01.002

Wakefield, E. M., Congdon, E. L., Novack, M. A., Goldin-Meadow, S., and James, K. H. (2019). Learning math by hand: The neural effects of gesture-based instruction in 8-year-old children. *Attent. Percept. Psychophys*. 81, 2343–2353. doi: 10.3758/s13414-019-01755-y

Wakefield, E. M., James, T. W., and James, K. H. (2013). Neural correlates of gesture processing across human development. *Cogn. Neuropsychol*. 30, 58–76. doi: 10.1080/02643294.2013.794777

Wakefield, E. M., Novack, M. A., Congdon, E. L., Franconeri, S., and Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci*. 21:e12664. doi: 10.1111/desc.12664

Wakefield, E. M., Novack, M. A., Congdon, E. L., and Howard, L. H. (2021). Individual differences in gesture interpretation predict children's propensity to pick a gesturer as a good informant. *J. Exp. Child Psychol*. 205:105069. doi: 10.1016/j.jecp.2020.105069

Watts, T. W., Duncan, G. J., Siegler, R. S., and Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educ. Res*. 43, 352–360. doi: 10.3102/0013189X14553660

Wiebe, S. A., Espy, K. A., and Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. latent structure. *Dev. Psychol*. 44, 575–587. doi: 10.1037/0012-1649.44.2.575

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev*. 9, 625–636. doi: 10.3758/BF03196322

Wu, Y. C., and Coulson, S. (2014a). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychol*. 153, 39–50. doi: 10.1016/j.actpsy.2014.09.002

Wu, Y. C., and Coulson, S. (2014b). A psychometric measure of working memory capacity for configured body movement. *PLoS ONE* 9:e84834. doi: 10.1371/journal.pone.0084834

Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., and van Lieshout, E. C. (2013). Individual differences in kindergarten math achievement: the integrative roles of approximation skills and working memory. *Learn. Individ. Differ*. 28, 119–129. doi: 10.1016/j.lindif.2013.09.012

Zheng, X., Swanson, H. L., and Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *J. Exp. Child Psychol*. 110, 481–498. doi: 10.1016/j.jecp.2011.06.001

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 April 2021
doi: 10.3389/fpsyg.2021.651725

Progressive Reduction of Iconic Gestures Contributes to School-Aged Children's Increased Word Production

Ulrich J. Mertens and Katharina J. Rohlfing*

Psycholinguistics, Faculty of Arts and Humanities, Paderborn University, Paderborn, Germany

The economic principle of communication, according to which successful communication can be reached by least effort, has been studied for verbal communication. With respect to nonverbal behavior, it implies that forms of iconic gestures change over the course of communication and become reduced in the sense of less pronounced. These changes and their effects on learning are currently unexplored in relevant literature. Addressing this research gap, we conducted a word learning study to test the effects of changing gestures on children's slow mapping. We applied a within-subject design and tested 51 children, aged 6.7 years (*SD* = 0.4), who learned unknown words from a story. The storyteller acted on the basis of two conditions: In one condition, in which half of the target words were presented, the story presentation was enhanced with progressively reduced iconic gestures (PRG); in the other condition, half of the target words were accompanied by fully executed iconic gestures (FEG). To ensure a reliable gesture presentation, children were exposed to a recorded person telling a story in both conditions. We tested the slow mapping effects on children's productive and receptive word knowledge three minutes as well as two to three days after being presented the story. The results suggest that children's production of the target words, but not their understanding thereof, was enhanced by PRG.

Keywords: word learning, child language acquisition, iconic gestures, reduction, economic principle of communication

OPEN ACCESS

Edited by:
Naomi Sweller,
Macquarie University, Australia

Reviewed by:
Paula Marentette,
University of Alberta, Canada
Suzanne Aussems,
University of Warwick,
United Kingdom

****Correspondence:***
Ulrich J. Mertens
ulrich.mertens@uni-paderborn.de

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

Received: *10 January 2021*
Accepted: *30 March 2021*
Published: *26 April 2021*

Citation:
Mertens UJ and Rohlfing KJ (2021) Progressive Reduction of Iconic Gestures Contributes to School-Aged Children's Increased Word Production. Front. Psychol. 12:651725. doi: 10.3389/fpsyg.2021.651725

INTRODUCTION

Reduction in Spoken Language and Gestures

How people structure information in speech depends on various factors, including what is assumed to be known, what kind of information is considered important, and what information the speaker wishes to focus on (e.g., Arnold et al., 2013). In this vein, studies on speech have shown that speakers exclude information when they tell a story for the second time to the same interlocutor and that stories told for the second time contain fewer details and fewer words (Galati and Brennan, 2010). Moreover, when referring to the same entity repeatedly, a speaker reduces the full lexical form by replacing it with a pronoun or a zero anaphora (e.g., Fowler et al., 1997; Galati and Brennan, 2010). Another form of reduction occurs when a word is produced less intelligibly (Bard et al., 2000, p. 2) by shortening its vocalization duration (Jescheniak and Levelt, 1994; Griffin and Bock, 1998;

Bell et al., 2009; Lam and Watson, 2010) and producing it without pitch accent (Gregory, 2002; Watson et al., 2008). Overall, these kinds of reductions occur during an interaction for predictable (Haspelmath, 2008) or already known referents. The advantage of using less information is a phenomenon already well studied and is related to the economic principle of communication (for an overview, see Arnold et al., 2013).

Similar to verbal behavior, gestures that encode the same referent vary in their quantitative and qualitative aspects to adapt to the listener's communicational needs (Gerwing and Bavelas, 2004; Galati and Brennan, 2014; Bohn et al., 2019) and in the interaction progress that contributes to emerging common knowledge (Clark, 1996). The similarity between verbal and gestural behavior is reflected in the current literature assuming that gesture and speech use the same communication planning processes (McNeill, 1992; Kendon, 2004). The two modalities function as one integrated system and are manifested in its temporal alignment (e.g., Jesse and Johnson, 2012; Esteve-Gibert and Prieto, 2014), in similar semantics (McNeill, 1992; Kendon, 2004), and in pragmatic aspects (e.g., Kelly et al., 1999).

Based on the well-acknowledged view that gesture and speech form an integrated system, in our study, we reasoned that speakers' gestures undergo similar changes as speech forms (Galati and Brennan, 2010, 2014). In this vein and focusing on iconic gestures, which are gestures that bear semantic information about objects and actions, Galati and Brennan (2014) showed that gestures become attenuated in size and iconic precision when produced for a known interlocutor compared to an unfamiliar interlocutor. Similar to lexical forms in Galati and Brennan (2010), shared knowledge was visible in gestures in the form of a reduction. Similarly, Gerwing and Bavelas (2005) revealed that with increased, mutually shared knowledge, gestures become physically more schematic while simultaneously becoming conceptually more complex. Whereas the dimensions of reduction are still largely unexplored (Koke, 2019), it seems that interlocutors with a certain degree of shared knowledge use less accurate gestures than interlocutors without shared knowledge (Gerwing and Bavelas, 2004). The latter type of interlocutors (without shared knowledge) displayed more elaborated, informative and precise gestures (Gerwing and Bavelas, 2004). Similarly, Jacobs and Garnham (2007) demonstrated an effect on adult participants' gestures that pertains to the interlocutors' established common ground: Gestures became less complex, precise, and informative when speakers communicated about toys with which listeners had also played. Along the same lines, Holler and Wilkin (2011) demonstrated that interlocutors, who talked about shapes on cards in order to sort them, mimicked each other's gestures during the dialog and that, as their mutually shared understanding increased, their gestures were produced less precisely. Overall, a reduction of gesture movements during an interaction and the loss of particular semantic aspects could be observed. It should be noted, however, that the reduction did not cause a loss of information in the context of the conversation. Instead, the relevant semantic information within the reduced gestures was available for the listener at any time because the listener could rely on the interaction history to link reduced gestures to referents introduced earlier on (Holler and Wilkin, 2011; Hoetjes et al., 2015).

In sum, the reviewed literature suggests that gesture production is adaptive to the listener's emerging knowledge. The body of research also supports cross-situational processing mechanism in memory: More specifically, an aggregation of features that seems to form an overreaching element that is used in a contextualization process. In this process, an ongoing event is interpreted in light of the emerging knowledge of the interlocutors. However, direct empirical evidence for the effects of adapted (i.e., reduced) gestures for learning is currently lacking.

Learning With Iconic Gestures

In contrast to the advantage of adapted gestures, gestural behavior itself is largely demonstrated to support language learning (see, e.g., Rohlfing, 2019 for a recent review). Several studies report an improvement in word learning for preschool children (e.g., Vogt and Kauschke, 2017a), elementary school students (e.g., Nooijer et al., 2014), and adults (e.g., Goodrich and Hudson Kam, 2009) in a word learning scenario in which iconic gestures accompany target words. However, most existing studies focus on younger children, thus, the evidence for older children is scarce (Rohlfing, 2019).

In the literature, two explanations are provided for the effectiveness of learning with gestures with regard to younger children. First, iconic gestures semantically enrich the encoding of unknown words (Capone Singleton, 2012) thus contributing to a long-term learning effect (McGregor et al., 2009), also referred to as slow mapping effect (e.g., Munro et al., 2012). In other words, new information is first processed in working memory (fast mapping) and then stored in long-term memory (slow mapping). According to word learning studies, the transition from working to long-term memory involves cognitive processes during sleep (Wojcik, 2013). These consolidation processes yield a memory trace that supports the retention of a novel word (e.g., Munro et al., 2012) and become visible as consolidation effect (for an overview, see Dudai, 2004; Wojcik, 2013). As already mentioned, in word learning, the contribution of iconic gestures was related to deeper processing: When a learner sees gestures performed, they evoke semantic elements that are not yet part of the word's mental representation (Kita et al., 2017). Consequently, binding a relation between the entity perceived (e.g., a practical action) and its abstracted features in the form of a gesture results in a richer internal representation that requires a deeper level of processing (Goldstone and Son, 2005; McNeil and Fyfe, 2012; Kita et al., 2017). In turn, a deeper level of processing seems to leave a greater memory trace (McGregor et al., 2009; Son et al., 2012). Other explanations for the beneficial effect of iconic gestures focus on gestures that are used by the learner. In these situations, the use of iconic gestures lightens the demands on the learner's working memory (e.g., Baddeley, 1986; Goldin-Meadow et al., 2001; Ping and Goldin-Meadow, 2010; Cook et al., 2012). For example, Goldin-Meadow et al. (2001) showed that children recalled a list of words better when they were allowed to gesture than when they were not.

Summarizing the existing research, Rohlfing (2019) points to the evidence that gestures support the learning of various word classes: nouns, verbs, and prepositions. While the acquisition of various word classes benefit from iconic gestures, the GSA framework, which is based on the idea that gestures arise from underlying motor or visual imagery, suggests that verbs require "complexive" attributes (Nomikou et al., 2019, p. 9) that might be better reflected in a multimodal way. This suggestion is grounded in empirical evidence that shows, for example, that children gesture more when describing a verb compared to a noun (Hostetter and Alibali, 2008, 2019; Lavelli and Majorano, 2016). Studies that investigated children's word acquisition support this finding by demonstrating that when children observe iconic gestures their verb learning benefits from this observation (e.g., Mumford and Kita, 2014; Aussems and Kita, 2020). Mumford and Kita (2014) argue that iconic gestures guide children's attention towards particular features of a scene which can enhance their semantic representation of unfamiliar verbs. Aussems and Kita (2020) demonstrated that iconic gestures foster the learning of locomotion verbs by preschool children. Yet another study showed that primary school children benefit from iconic gestures when learning locomotion verbs, but no enhanced learning effect was observed for object manipulation and abstract verbs (Nooijer et al., 2014). This finding indicates that iconic gestures' influence on verb learning varies between verb categories.

Both explanations—to enrich the encoding semantically and to lighten working memory—that regard the facilitative effect of iconic gestures on word learning account for the effect that a single gesture has during a learning experience. We now turn to the questions of how and in what manner multiple presentations of a gesture can enrich the encoding of words.

Learning With Variations of Gestures

To our knowledge, variation in iconic gestures has not been considered in word learning studies to date. Although the phenomenon of reduced gestures seems natural, it has not been studied systematically during learning situations. When gestures were used to support word learning in previous studies, they remained unchanged even when presented several times. In these studies, when the gesture consistency was an issue, it was achieved by presenting participants with gestures of video-recorded persons (e.g., Vogt and Kauschke, 2017a) or programmed social robots (e.g., Vogt et al., 2017). In contrast to gesture consistency, few studies tackled the issue of gesture reduction. Variation in gestures can be achieved in manifold ways and can occur in all gesture phases: preparation of the gesture, in which the hand starts to move from a resting position, the stroke, when a peak in movement is performed, and the retraction phase, in which the hand(s) switch to a rest position or to another gesture (Kendon, 1972, 1980; see for overview: Wagner et al., 2014).

One possibility to vary a gesture is to provide different aspects that refer to a specific referent. This is particularly relevant for iconic gestures that convey semantic information through their form (as in McGregor et al., 2009). For example, showing how an object falls could be depicted in a reduced iconic gesture by a quick hand movement that uses a downward movement. This event could also be depicted with an even more reduced gesture using only one finger. In contrast, the full gesture could involve an arm movement to depict the length of the downward movement, while the hand would additionally depict semantic features of the object.

It has been observed that such a reduction occurs naturally when speakers repeatedly refer to the same referent. They usually reduce some properties of the gesture without changing or losing the core meaning of the gesture (Gerwing and Bavelas, 2004; Holler and Wilkin, 2009, 2011; Galati and Brennan, 2014; Hoetjes et al., 2015; Bohn et al., 2019). As already stated above, the reduction of gestural presentation is not only a byproduct of emerging common knowledge: When the form properties change, the semantic information of the gesture changes as well. In the following paragraphs, we present arguments for why progressively reduced gestures, rather than gestures that are presented in the same manner, might improve learning.

First, learners aggregate information across different experiences with a novel referent and its labeling to discover the relevant properties and features (Yu and Smith, 2007). Following evidence provided above suggesting that gestures contribute to the semantic encoding (McGregor et al., 2009; Capone Singleton, 2012; Vogt and Kauschke, 2017b), we assume that gestures are part of semantic knowledge that is generated during exposure and will be used for learning. We further reason that children's semantic knowledge is even more enhanced when learners experience different versions of a gesture because different semantic features of the referent are embodied in each version. In addition, these semantic features become contextualized in the process of unfolding knowledge and might become conceptually more complex with each version (Gerwing and Bavelas, 2004). This contextualization process might require more cognitive effort from the learner to bind the different features in the sum as relevant for the referent. To put it in other words, each time the gesture is performed to supplement an unknown word, it will provide additional, relevant information that needs to be related to the word. This is because the gesture becomes more and more abstracted from the referent.

We argue that this contextualization, namely, to relate the abstracted (or reduced) content to the referent, is an effort that fosters a deeper memory trace. In a similar vein, Son et al. (2012) studied under what situational circumstances children generate relational information that leads to generalization across trials. They concluded that for a word to become generalized, there should not be too much concrete information involved in the labeling experience (Son et al., 2012, p. 9). When learning instances are too specific, this experience might activate only an immediate memory system and not generate any relational information. This work led us to hypothesize that the interpretation of several reduced features accumulated in gesture results in meaningful relationships between the depicted features and the concrete referent and, furthermore, contributes to children's robust word learning.

Further support for our premise comes from research that shows that movements in the field of view have a distracting effect and can interfere with the participants' task performance

(Lavie, 1995, 2005; Rees et al., 1997; Forster and Lavie, 2008). Distractors that are unrelated to a task and only appear in the periphery attract participants' attention when cognitive resources are available. More importantly, task-relevant distractors are just as likely to interfere with task performance as irrelevant distractors (Forster and Lavie, 2008). When applying Lavie's attention theory to children who observe fully executed iconic gestures constantly, we derive the idea that these children pay attention to such gestures (Kelly et al., 2010; Wakefield et al., 2018a); however, seeing fully executed gestures multiple times might have a distracting effect (Forster and Lavie, 2008). Using our earlier example, a gesture that depicts the event of a falling object can be performed by raising the hand above the head and then moving the hand in a quick motion toward the floor or by a short movement with only one finger. As illustrated above, an interlocutor can gather the full meaning of a reduced gesture when it is performed in context. Paying attention to a fully executed gesture requires cognitive resources that are not directed to the accompanying word. This assumption is supported in studies showing that higher cognitive load is reflected in participants observing movements and solving linguistic tasks (Rees et al., 1997). In contrast, when observing a reduced gesture, a child might focus more on the accompanying word. As such, experiencing a reduced gesture depicting the event of falling down might distribute children's attention more equally on the gesture as well as the word. Consequently, a rich memory of the referent can be created because cognitive resources are distributed more economically to build better-balanced relational structures between the semantic features in the gesture and the label.

In sum, our assumption that progressively reduced iconic gestures might foster a memory trace of an unknown word is based on the following: Their reduced movements (i) require a contextualization that let a relational structure between the word and the reduced features of the gestures emerge through aggregation of semantic features and (ii) are less distracting and can even create a processing focus on the label over time. The first premise pertains to cognitive learning mechanisms that appear to be activated during the observation of iconic gestures. For the second premise, we have argued that learning becomes enhanced due to more balanced distributed cognitive resources when observing progressively reduced iconic gestures. Together with the above-mentioned fact that reduction occurs in natural communication, these premises provide a basis for our study.

The main goal in our study was to investigate whether children are sensitive to reduced iconic gestures and whether their long-term word learning (production and reception) is enhanced when observing progressively reduced iconic gestures. Whereas the existing body of research focuses on preschool children (Rohlfing, 2019), we investigated older children to extend an existing body of research to which we can associate our study with respect to both advantages of (i) gestural presentation for unknown words as well as (ii) long-term memory. Studies have shown more potent effects for children when tested with delay to initial exposure to a target word (e.g., Munro et al., 2012). Furthermore, being aware that word learning comprises the acquisition of many word classes, our aim was to account for this diversity in our study design, for which we used nouns and verbs as target words.

MATERIALS AND METHODS

Participants

Fifty-one first graders, including 25 females and 26 males from two schools in the region of Meerbusch (North-Rhine Westphalia) in Germany, participated in this study. The participants ranged in age from 6.0 to 7.4 years (*mean* = 6.7; *SD* = 0.4). Socioeconomic status data were not collected from children, but the population from which the sample was drawn was predominantly middle to upper-middle class.

Stimuli

For our study, we used a word learning setting in which words are embedded within a story—a previously designed successful method for children (Nachtigäller et al., 2013; Vogt and Kauschke, 2017a). The story, target words, pictures, and iconic gestures were used from another word learning study (Vogt and Kauschke, 2017a,b). In their study, Vogt and Kauschke (2017a,b) demonstrated that preschoolers gained greater word knowledge when a speaker accompanied words with iconic gestures compared with attention gestures. For our study, we modified the story in terms of the frequency and the number of target words. In total, we embedded 4 nouns and 4 verbs within the story, and each of them occurred three times. Whereas the four nouns referred to animals, the verbs referred to locomotion. The eight words were German words chosen by Vogt and Kauschke (2017a) and identified as low-frequency words (University of Leipzig, 1998–2013). In support of this, four- and five-year-old German children (*n* = 16) were asked to name the stimuli, and none of the children could name any of the stimuli (Vogt and Kauschke, 2017a). We supplemented the ratings by asking adults (n = 10) to name the stimuli. Only one of the ten adults was able to name one word (a noun).

As in the original study, children watched a recorded person who told the story and accompanied the target words with gestures (Vogt and Kauschke, 2017a). Additionally, we extended the multimodal presentation of the target words by presenting reduced gesture versions (see **Supplementary Material**). To ensure consistent word pronunciation of the target words, we desynchronized gestures from the spoken word by performing the gestures shortly after the spoken word. This way, the stroke of the gesture was not synchronous with the target word. Instead, the gesture was presented right after the word was produced.

We created two reduced gesture versions for each gesture. With every reduction, a gesture becomes less complex and less precise (Jacobs and Garnham, 2007) by lowering the gesture's level of detail and shortening its trajectory (Galati and Brennan, 2014; Hoetjes et al., 2015). Reduced gestures for nouns and verbs were achieved by indicating the shape of an object and/or the action movement with less accurate spatial information about the referent's location. For both word classes, this reduction led to a shortened duration of the gesture phases. In **Figure 1**, the

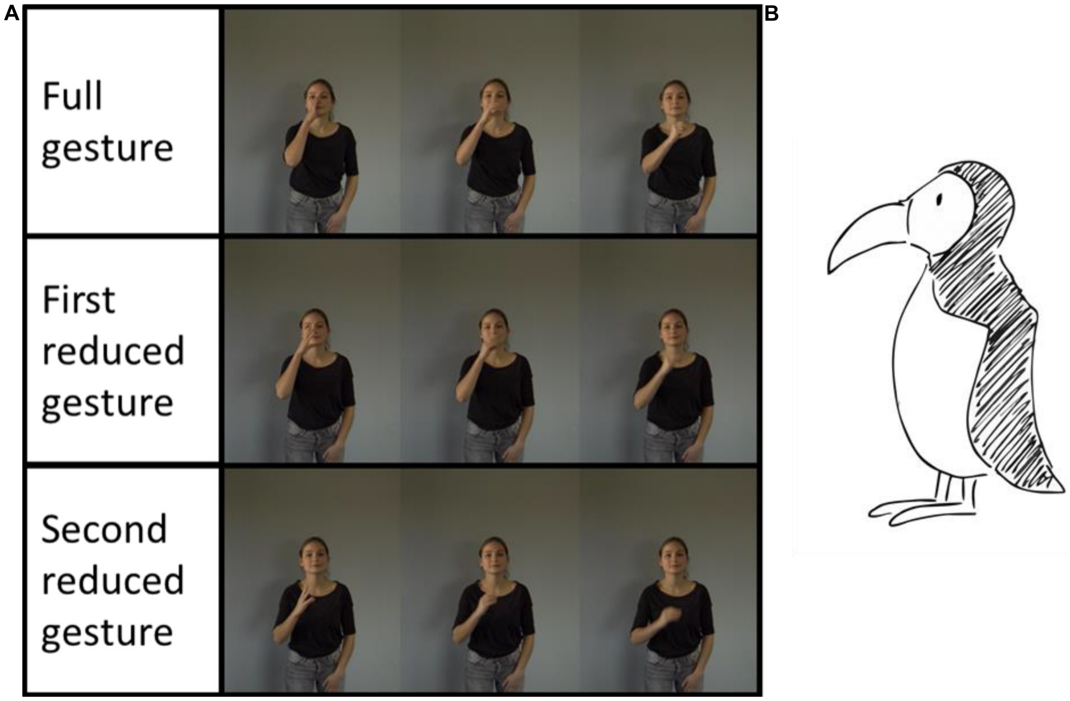


FIGURE 1 | (A) The three gesture versions of the iconic gesture for the noun "auk". The first row depicts the fully executed gesture; the second row depicts the first reduced gesture version; the third row depicts the second reduced gesture version. **(B)** The picture displays the referent "auk" (Copyright © 2013 Joy Katzmarzik leap4joy graphics; reprinted with permission).

fully executed gesture accurately depicts the shape and location of the auks' beak, whereas the second reduced gesture only implies the beaks shape and its spatial location. Similar characteristics account for reduced gestures of locomotion verbs. For example, the fully executed gesture for "to creep" depicts the referent's movements and a clear horizontal movement direction. The reduced gesture version indicates only a horizontal direction with the speaker performing an almost arcuate hand movement from the left to the right. For reduced gestures of both word classes, the stroke phase is not clearly separable from the preparation and retraction phases.

Design and Procedure

For our investigation, we visited children at their respective schools for two sessions. We selected five different classes from two schools. Before starting the first session, the experimenter visited the children in their classroom to introduce himself and the project. The children's parents were informed and asked for their consent by letter. The study commenced after parents provided written consent to their children's participation, which is in accordance with Paderborn University's ethics procedures for research with children. The procedure and consent forms were approved by the university's ethical committee. The children also provided verbal consent before participating. Additionally, they were informed that they could discontinue the interaction at any time.

The two sessions for our investigation took place in a one-to-one constellation with only the child and the experimenter in the room. In both sessions, a child sat down in front of a monitor set up on a table. The experimenter was sitting at another table opposite the child. A plexiglass panel was placed between the tables as a precautionary measure due to the Coronavirus pandemic (see **Figure 2**). The first session lasted approximately fifteen minutes and the second session about five minutes. The children's responses during the testing were videorecorded for later analysis. The experimenter was aware of the purposes and hypotheses of the study but blind to the gesture condition that a child experienced.

Learning: After a short chat about how the children feel being in first grade, the experimenter explained to the children that he wanted to show them a video of a young adult who would tell a story about her first-grade experience. After the child's consent, the experimenter started the video.

In the video, a woman told a story about animals and actions (that served as target words). We applied a within-subject design: To identify the effect of gesture reduction on children's slow mapping of novel words, half of the iconic gestures became progressively reduced. For this, children were exposed to three versions of gestures that appeared progressively reduced. Furthermore, the story was designed for each target word to occur three times in succession, without other target words being mentioned. During this part, a picture with the referent was shown next to the speaker (see **Figure 3**). Showing children an image of a referent within the experimental setup is necessary for testing children's word knowledge that was administered after the learning phase.

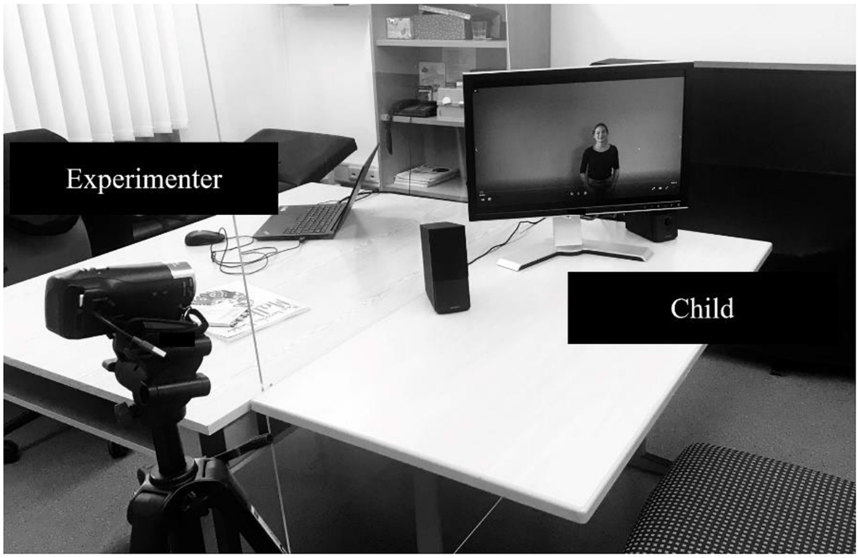


FIGURE 2 | Study setup.

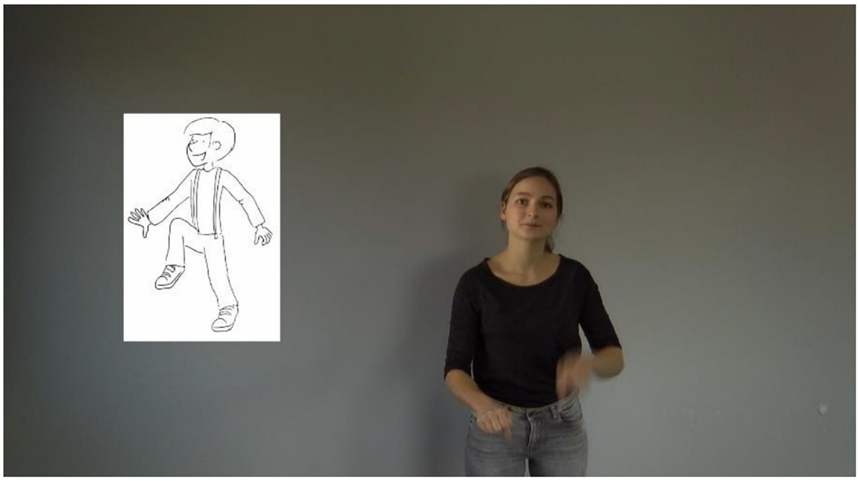


FIGURE 3 | Recorded storyteller performing the gesture for the German target verb "*staksen"* [to lift the legs alternately]. Next to her, the referent appears as an image (Copyright © 2013 Joy Katzmarzik leap4joy graphics; reprinted with permission).

In our pilot study, when we put eight target words that the children had to remember in one story, we obtained floor learning effects. Our interpretation was that recalling eight target words might have overwhelmed the children. Our attempts to reduce the load were successful, and we found that children performed better when they watched the story in two parts. For this reason, we first presented one part of the story (with four target words) and tested children's learning performance after a break of three minutes. After testing children's word knowledge, we continued with the second part of the story (with different four target words) that was followed three minutes later by a second test. According to this study design, children's receptive and productive knowledge of the target words was assessed twice, once after each part (see **Figure 4**). This design raises the issue that children might be aware of the story's purpose during the second part. Consequently, children might learn target words from the second part better. To avoid this bias, we created two story versions in which the target words were embedded differently. The target words that occur in the first part of the first story version were embedded in the second part of the second story version and the target words that occur in the second part of the first story version were embedded in the first part of the second story version. Every part contained four different target words (two nouns and two verbs). Furthermore, each story version was created in two ways, depending on whether target words were accompanied with progressively reduced iconic gestures (PRG), or fully executed iconic gestures (FEG). In total, we created four videos that

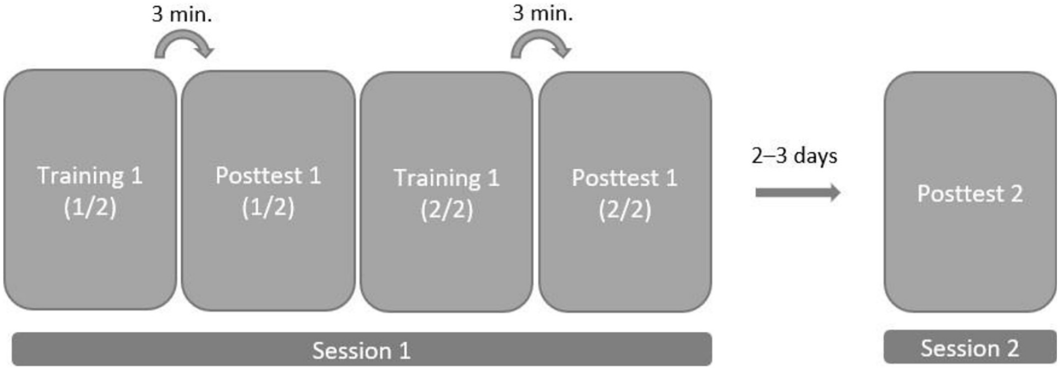


FIGURE 4 | The study design: In the first session, the story was split into two parts. Each training part was followed by a posttest that was administered after a 3-minute delay. Posttest 2 took place during the second session.

differed in the order of the target words and the gesture versions (progressively reduced or fully executed) that accompanied the target word.

Testing: The main goal of this study was to identify the long-term effects of reduced gestures on children's productive and receptive word knowledge. For this purpose, children's slow mapping performance was assessed three minutes after hearing each part of the story. During the break between training and testing, the children were asked to color a picture. A further long-term effect on children's receptive and productive knowledge was tested in the second session that took place two to three days after the first session. During the testing sessions, children were shown pictures of the target words, similar to those shown in the video. However, for the nouns, the pictures displayed the animals from a different perspective, and the verb pictures featured a girl instead of a boy. Throughout the testing, the experimenter provided neutral feedback to the children's answers.

As mentioned above, our testing assessed children's performance in word production and understanding. During the production task, the experimenter asked the child: "Can you tell me what kind of action the girl in the picture is performing?" or "Can you tell me what kind of animal is shown in the picture?" At the same time, the picture of the referent was shown on the monitor (see **Figure 5**). In the case when children did not provide an answer within five seconds after the question was raised, the experimenter asked the children if they had any idea. If another five seconds elapsed without an answer, the experimenter moved on to the next picture and said "no problem, let's look at the next picture" or provided a similar form of reassurance. The experimenter also continued with the next picture when the children gave a correct or incorrect answer or made it explicit that they did not know the answer. In that case, the experimenter said, for example, "let's look at the next picture."

Children's performance was scored according to a coding system. Children obtained (i) two points when both the onset and the offset of the word were correct and they used the correct number of syllables, (ii) one point when they produced either the onset or the offset of the word correctly and used the correct number of syllables, and (iii) zero points when they produced the word onset and the offset incorrectly or when the number of syllables was incorrect.

Fifteen percent of production responses were randomly selected and coded by an independent research assistant. We

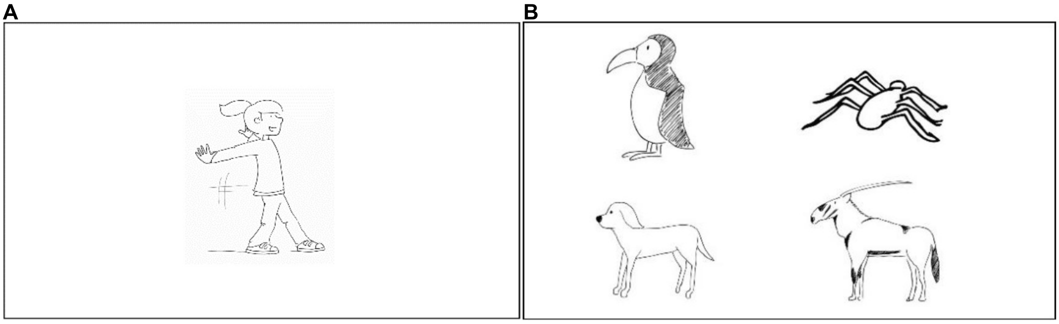


FIGURE 5 | Monitor screen during **(A)** the production task and **(B)** the reception task. In the example for **(A)**, the girl is performing the action "to slide." In the example for **(B)**, at the bottom right, the referent for the target word "beisa" is displayed. A distractor object that looks similar to the target referent is presented to the left of the "beisa". In the first row on the left, another target word, "auk," is presented as a distractor. Next to the "auk," a random referent is shown as an additional distractor (Copyright © 2013 Joy Katzmarzik leap4joy graphics; reprinted with permission).

TABLE 1 | Children's mean production scores (SD) in the testing Session 1 (T1) and testing Session 2 (T2).

	Production			
	T1		T2	
	PR	CF	PR	CF
Words	1.92 *(1.77)*	1.16 *(1.47)*	2.59 *(1.91)*	1.20 *(1.48)*
Nouns	1.08 *(1.23)*	0.75 *(1.26)*	1.24 *(1.35)*	0.67 *(1.10)*
Verbs	0.84 *(1.22)*	0.41 *(.75)*	1.35 *(1.39)*	0.33 *(1.33)*

The maximum word score was 8 points with 4 points for nouns and verbs each. PRG = progressively reduced iconic gestures; FEG = full executed iconic gestures.

measured interrater reliability using Cohen's Kappa (Cohen, 1960) and obtained an agreement of $k = 0.92$.

In the reception task, children were presented the target referent with three distractors; all referents formed a 2x2 arrangement (see **Figure 5**). The probability of choosing the correct answer by chance was 25%. The distractors in the arrangement included a picture similar to the target referent, another target picture out of our study (same word type), and a random picture. The testing started by asking the child, for example, "Can you touch the picture where you see the beisa?" When children did not point at the screen within five seconds of being asked the question, the experimenter asked again if they could point at the screen. If another five seconds elapsed without an answer, the experimenter moved on to the next referent and said to the child "It doesn't matter, let's look at the next picture" or provided a similar form of reassurance. The experimenter also continued to the next referent when the children pointed at the screen or made it explicit that they did not know the answer. The experimenter initiated this progression with words such as "let's look at the next picture!" After testing session 2, each child's performance was scored according to a coding system: Children obtained one point for each correct answer and zero points for an incorrect answer.

Data Analysis

We applied an omnibus 3-way analysis with the independent variables gesture (progressively reduced iconic gestures (PRG), fully executed iconic gestures (FEG)) and time (T1, T2) for testing effects on nouns and verbs for both production and reception. Greenhouse–Geisser corrections were applied where necessary. Significant interaction effects were resolved by Bonferroni corrected *post hoc* pairwise comparisons. For the production task, we additionally conducted an item analysis. We first report on the production task before and then turning to the reception task.

RESULTS

Word Production

Children's performance was measured on a scale from 0 to 16 for word learning (8 points for words accompanied by progressively reduced gestures (PRG) and 8 points for words accompanied by fully executed gestures (FEG)). Children achieved a mean of 3.17 points (SD = 2.57; range: 0–10) during testing Session 1. During the testing Session 2 children achieved a mean of 3.80 points (SD = 2.58; range: 0–12). Their performance is displayed in **Table 1**.

The ANOVA confirmed an intermediate significant interaction effect gesture × time ($F(1,50) = 5.55$, $p < 0.05$, $\eta p^2 = 0.10$), reflecting that children scored differently in the gestural conditions and that the difference between conditions depended on the time of retention. In *post hoc* analyses, multiple pairwise comparisons revealed that children achieved higher scores in Session 1 when words were presented with PRG than when words were accompanied with FEG ($p < 0.05$). Similarly, in Session 2, children scored higher in the PRG than in the FEG condition ($p < 0.01$). These results suggest that children's word production was enhanced when they were exposed to presentation with PRG. Further analyses revealed that in the PRG condition, children achieved a higher score during T2 than during T1 ($p < 0.05$) suggesting that the effect of PRG became more pronounced over time (see **Figure 6**). For FEG, the *post hoc* analysis revealed no differences between T2 and T1 ($p = 0.86$). The ANOVA yielded no further significant interaction effect for gesture × word class × time ($F(1,50) = 1.56$, $p = 0.22$, $\eta p^2 = 0.03$), word class × time ($F(1,50) = 0.67$, $p = 0.42$, $\eta \mathrm{p}^2 = 0.01$) or gesture × word class ($F(1,50) = 1.05$, $p = 0.31$, $\eta p^2 = 0.02$) indicating that nouns and verbs were produced similarly at both points in time and under both gesture conditions.

In the next step, we applied an item analysis to assess the item's quality within the FEG and the PRG condition. The item difficulty ranges from 0.06 to 0.38 indicating that producing the target words can be considered as quite difficult for the participating children. **Table 2** shows that the frequency distribution of item difficulty is lower for seven out of eight items within the PRG condition. The item "fennec", however, was an exception, because it similarly difficult in both conditions. This analysis confirms that most items were learned more easily within the PRG conditions.

Reception Task

Children could score 8 points in the reception task (4 points for words accompanied by PRG and 4 points for words accompanied by FEG). In testing Session 1, children obtained a mean of 6.41 points (SD = 2.00; range: 0–8 points). During the testing in Session 2, children achieved a mean of 6.70 points (SD = 1.84; range: 2–8 points). Differentiating between word types (nouns and verbs), children could achieve 4 points for each word type (2 points for words accompanied by PRG and 2 points for words accompanied by FEG). For nouns, children obtained a mean of 3.08 points (SD = 0.73 ranging from 0–4) in testing Session 1. During testing in Session 2, children achieved a mean of 3.32 points (SD = 1.27 ranging from 0–4 points). With respect to verb reception, children obtained a mean of 3.36 points (SD = 0.65 ranging from 0–4) in testing Session 1. During testing Session 2, children achieved a mean of 3.42 points (SD = 0.99 ranging from 1–4). The probability to choose the correct answer by chance was at 25% within the reception task. With children's responses being at 80% in testing Session 1 and 83% in testing Session 2 for words in general but also 77 % in testing Session 1 and 83% in testing

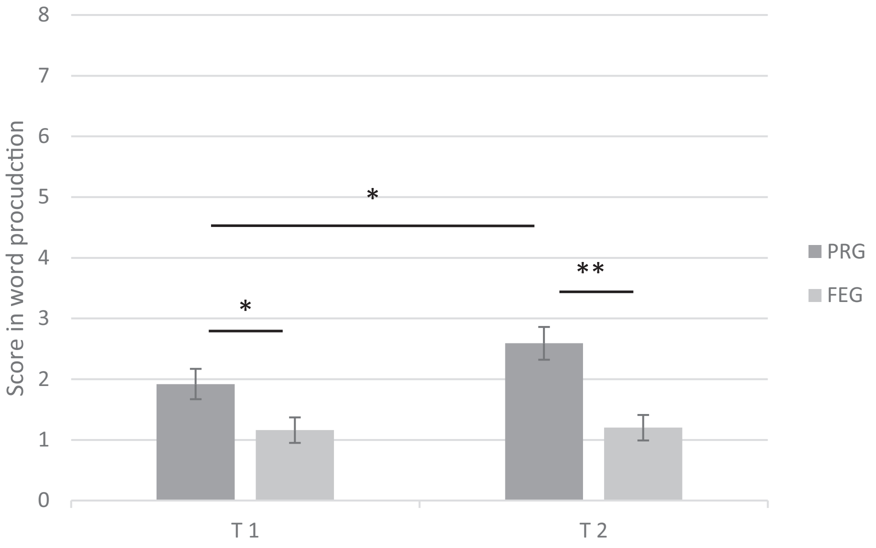


FIGURE 6 | Children's word production score (*SE*) at the first (T1) and second (T2) testing. PRG = progressively reduced iconic gestures; FEG = fully executed iconic gestures; children could score 8 points in both conditions, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 2 | Mean score, standard deviation (*SD*), and difficulty for each item (target word) in the PRG and FEG condition (PRG = progressively reduced iconic gestures; FEG = fully executed iconic gestures).

Item	**Ralle** ***"rail"***		**Alk** ***"auk"***		**Fennek** ***"fennec"***		**Beisa** ***"beisa"***	
Condition	**FEG**	**PRG**	**FEG**	**PRG**	**FEG**	**PRG**	**FEG**	**PRG**
Mean (*SD*)	0.32 (*0.73*)	0.76 (*0.95*)	0.16 (*.54*)	0.48 (*0.80*)	0.40 (*0.78*)	0.44 (*0.81*)	0.41 (*0.81*)	0.85 (*0.91*)
Item Difficultiy	0.16	0.38	0.08	0.24	0.20	0.22	0.21	0.42
Item	**Gliddern** ***"to slide"***		**Staksen** ***"to lift the legs alternately"***		**Krauchen** ***"to creep"***		**Retschen** ***"to slide backwards"***	
Condition	**FEG**	**PRG**	**FEG**	**PRG**	**FEG**	**PRG**	**FEG**	**PRG**
Mean (*SD*)	0.42 (*0.83*)	0.64 (*0.92*)	0.26 (*0.62*)	0.62 (*0.89*)	0.13 (*0.38*)	0.37 (*0.71*)	0.22 (*0.60*)	0.48 (*0.81*)
Item Difficultiy	0.21	0.32	0.13	0.30	0.06	0.18	0.11	0.24

Session 2 for nouns and 84% in testing Session 1 and 86% in testing Session 2 for verbs in specific, we can state that children performance in word reception was well beyond the chance level.

The ANOVA revealed no significant interactions, gesture × word class × time ($F(1,50) = 1.73$, $p = 0.20$, $\eta p^2 = 0.03$), word class × time ($F(1,50) = 0.93$, $p = 0.34$, $\eta p^2 = 0.02$), gesture × word class ($F(1,50) < 0.01$, $p = 0.93$, $\eta p^2 < 0.01$), gesture × time ($F(1,50) = 0.01$, $p = 0.92$, $\eta p^2 < 0.01$), revealing that the children's performance in the reception task seems robust against the gesture presentation and time condition for nouns as well as verbs (see **Table 3**).

TABLE 3 | Children's mean reception scores (SD) in testing Session 1 (T1) and testing Session 2 (T2).

	Reception			
	T1		**T2**	
	PRG	**FEG**	**PRG**	**FEG**
Words	3.31 *(0.99)*	3.10 *(1.01)*	3.41 *(0.78)*	3.29 *(1.06)*
Nouns	1.55 *(0.64)*	1.53 *(0.09)*	1.73 *(0.57)*	1.59 *(0.70)*
Verbs	1.75 *(0.05)*	1.61 *(0.60)*	1.73 *(0.45)*	1.69 *(0.54)*

The maximum of word score is 4 points with 2 points for nouns and verbs each. PRG = progressively reduced iconic gestures; FEG = fully executed iconic gestures.

DISCUSSION

Whereas the economic principle of communication is well studied for verbal communication, little is known about means and effects of economic communication in gestural behavior. Aiming to close this gap, our study was designed to experimentally investigate the effects of progressively reduced iconic gestures (PRG) on children's word learning at a mean age of 6.7 years (*SD* = .4). More specifically, we asked whether children's slow mapping can be enhanced by presenting PRG in contrast to consistently fully executed iconic gestures (FEG). This new form of gestural presentation was motivated by two research strands: One strand includes studies demonstrating that iconic gestures comprise reductions of the referent's semantic features (e.g., Kita et al., 2017). Along these lines, we reasoned that this reduction leads to a more abstracted presentation of the referent, which is important to induce deeper memory processing resulting in a better learning outcome (Mumford and Kita, 2014; Son et al., 2012). Additionally, our study was motivated by the finding that common ground between interlocutors affects their gesture performance in a way that their gestures become reduced, but the reduction causes no loss of information in the context of the conversation (e.g., Holler and Wilkin, 2011; Galati and Brennan, 2014; Hoetjes et al., 2015). We reasoned that repeatedly observing FEG can lead to distracting effects, whereas through PRG, cognitive resources are distributed more economically and thus better balanced for processing meaningful input from a gesture and its accompanied label (Forster and Lavie, 2008; Kelly et al., 2010). Combining these two research strands, we expected children to retain target words accompanied by PRG better than words accompanied by FEG.

In our study, children were presented with eight unknown words: four nouns and four verbs. The unknown words were embedded in a story. Applying a within-subject design, children received four target words presented by PRG and four other target words presented by FEG. All children participated in both conditions. Our analysis focused primarily on long-term effects because retaining a word for several minutes or several days indicates that the word has been acquired robustly (e.g., Munro et al., 2012; Wojcik, 2013). For this reason, children's performance in word reception and production were assessed at two different points in time: after a delay of three minutes and after two to three days.

For word reception, we found no significant effect, neither when looking at the differences between the presentations nor when looking at what point in time the assessments occurred. We can therefore conclude that the reception of unknown words seems robust to our experimental manipulation. Furthermore and because of the high scores obtained in both conditions, our results suggest that first graders are generally strong in word reception. The referent's picture might have been a beneficial (nonverbal) resource for formulating the correct answer. Thus, it seems reasonable that older children are experienced enough to recall a word meaning with the presentation of a picture's referent—even if it is displayed from a different perspective. In contrast to our results, strong long-term effects on word reception were reported for younger children at the age of two, when the learning process was supported by iconic gestures (Horst and Samuelson, 2008; McGregor et al., 2009; Munro et al., 2012). It seems likely that the word reception test in our study was too easy for the children, which is a limitation of our design. In future studies, it would be more appropriate to design a testing procedure that requires the reception to be embedded in more demanding tasks, such as the understanding of text that contains the target words.

Regarding word production, we found that children were able to learn target words accompanied by PRG more successfully than words accompanied by FEG. In accordance with previous studies that revealed long-term effects of learning with gestures (McGregor et al., 2009; Munro et al., 2012), we found that the advantage of the PRG presentation was more pronounced when tested two to three days after initial exposure. We explain this as being a result of children's greater sensitivity to a word's presentations accompanied by PRG because the children experienced various forms of the gesture that might have fostered rich word concepts. These concepts were then available for the children during the assessment of their word production performance. The concept richness might be due to a greater variation in semantic properties in PRG, which are all related to each other. For example, the fully extended gesture for "to creep" contains several finger movements and a long horizontal trajectory, while the second reduced form contains no finger movements and only a short, almost arched trajectory. By removing semantic aspects from an iconic gesture, children might focus on the remaining semantic aspects from the reduced gesture. This way, children are exposed to a broader spectrum of semantic aspects within gestures that allows them to build a more substantial memory trace. In this form of gesture support, the variety of gestures includes a higher level of multimodal information. Thus, children can build up their semantic knowledge by continuously picking up semantic features that are novel or incongruent with their current word conception. This selected and contextualized exposure to various semantic features fosters the process of elaborating an existing representation and leads to a broader relational knowledge of the referent event. In support of this explanation, much research has emphasized that sematic knowledge drives the successful retrieval of a word's label for production (e.g., McGregor et al., 2002; Capone and McGregor, 2005; Capone Singleton, 2012).

While variations in gesture lead to a more complete and distinct representation in memory, it should be noted that the presentation of PRG included consistency in the presentation of the target word. This way, in repetitions of the presentation, the word became the invariant element (Son et al., 2012). Consequently, the word likely became a focus leading to a stronger memory trace by serving as a strong link between the semantic features within the gesture versions and the label. We argue that this focus also accounts for the beneficial effect of the PRG presentation that leads to stronger word production performance in a long-term. Son et al. (2012) have demonstrated that when cognitive effort is intensified to interpret perceptual events in the context of a word, a stronger relation between the label and the referent is created. The cognitive effortful processes that include extracting, supplementing, and contextualizing semantic features from PRG is likely to provide the semantic link that is needed to retain and recall a word in the long-term (Capone Singleton, 2012).

Experiencing RPG can clearly be viewed as contextualization that is taking place with regard to the ongoing gain of knowledge that the child is experiencing. However, it is important to note that following this explanation, it might also be possible that children's learning would benefit from presenting words with gestures that are not reduced but are instead presented each time differently. Further studies need to account for this alternative explanation. In line with our argumentation highlighting the relevance of semantic features in the facilitation process, we hypothesize that three unrelated gestures will not have the same beneficial effect on the production of unknown words.

As discussed above, our study demonstrates that children's slow mapping was enhanced when they were exposed to PRG gestures. To identify if specific stimuli drive this finding, we compared how well children learned each word in the PRG and the FEG conditions. The analysis revealed that all words, except the noun "fennec", were easier to produce when children observed PRG. Producing the word "fennec" appears to be equally challenging within the PRG and the FEG conditions. Interestingly, the gesture versions for "fennec" are executed with no movements within the stroke phase (the phase that contains the maximum semantic information density). All other gestures included movements within the stroke phase. We suggest that reducing a gesture that is void of movement in the stroke phase generates a lower variety of semantic features and can be interpreted effortlessly. The lower variety of semantic features, which seems to be easily processed, does not appear to contribute to the current internal word representation. The iconic gesture for "fennec" depicted the large ears of the animal. While the fully executed gesture version depicted the ears at an appropriate position on the head, the reduced gesture versions depicted the ears at less accurate positions. The reduced iconic gesture versions of other referents, like the peak of the auk, were reduced more strongly, involving a reduction of both the object (the peak) and the spatial position (see **Figure 1**). However, it also stands to reason that the item difficulty for "fennec" is similar in both conditions because it was simply not sufficiently reduced and not because of the missing movements within the stroke phase.

OUTLOOK

Our study indicates that PRG enhanced children's long-term word production in general, but no differences in learning nouns versus verbs were found. These findings are somewhat surprising considering that literature points out that the acquisition of verbs requires more complexive attributes than nouns (Nomikou et al., 2019). While nouns can be drawn from relatively established referential frames, verbs refer to events that are complex and less transparent to single out concrete semantic features (e.g., Gentner, 1982; Hirsh-Pasek and Golinkoff, 2006; Heller and Rohlfing, 2017). In this vein, other studies suggest the possibility that the acquisition of verbs benefits from multimodal presentations comprising additional semantic features (Goodrich and Hudson Kam, 2009; Kita et al., 2017; Wakefield et al., 2018b). For example, Mumford and Kita (2014) argue that extracting relevant features is one of the key elements in verb learning. With this in mind, it would be reasonable to expected that verbs were better learned due to the broader variety of semantic features within the PRG condition. While the omnibus 3-way ANOVA does not confirm significant effects between the word classes, a descriptive level of analysis shows that the studied children learned verbs accompanied with PRG as well as they learned nouns accompanied by PRG. In contrast, the children learned only half as many verbs as nouns when both word classes were accompanied with FEG. This descriptive analysis indicates the possibility that with increased power, various forms of gestures might be a method that responds better to demands in verb acquisition. Further research is needed to investigate whether PRG are particularly conducive to the acquisition of verbs.

Our second premise outlined in the introduction is that the movements themselves also play a role in learning with PRG. We have argued that children's production of novel words becomes enhanced with PRG because children can focus more on the label provided. While we found enhanced word learning effects in the long term, we did not investigate how different gesture conditions influenced children's attention. Future research can thus follow up an investigate how different iconic gesture versions affect children's attention.

LIMITATIONS

As mentioned above, our study has some limitations. First, we have argued that reduction in gestures can enrich children's semantic word knowledge by enabling a deeper encoding process induced by the reduced movement processing. It remains an open question whether the use of different iconic gestures would result in a similar learning effect.

Another limitation is the fact that children performed poorly in the production task, whereas they reached high scores in word reception. It seems reasonable to assume that the children's production scores would have been higher if the target words had been presented more frequently. However, the PRG condition required us to reduce each gesture only twice, to ensure the reductions between the different versions were noticeable. Consequently, the occurrence of each target word was limited to three times.

Finally, we decided to desynchronize the presentation of the spoken word from its accompanying iconic gesture. This was necessary to ensure that the presentation of the word was the same in each repetition. Normally, words are produced simultaneously with gestures. Consequently, as a gesture is reduced, the accompanying word's phonological form is also reduced. Since this confounds the effects of word with gesture presentation, we attempted to design our study so that it would avoid this problem. The desynchronization of the gesture and the word might have had an effect on children's learning outcome, as it seems easier for children to pay sufficient attention to a gesture and the target word. One way to perform variations of gestures simultaneously with the target word would be to use a social robot as storyteller. Despite a small sample size, this concept has shown promise in positively influencing word learning with PRG in preschool children (Mertens, 2017).

SUMMARY AND CONCLUSION

With our study, we have demonstrated that children's long-term word learning becomes enhanced through exposure to progressively reduced iconic gestures (PRG). The novelty of our research resides in the systematic description and experimental investigations of gestures that vary in their form when repeated during a word learning scenario. We have demonstrated the effects of PRG on productive word learning and offered thorough explanations. Our findings contribute to the growing evidence that a key element in supporting long-term learning processes is to reduce the learning content during its visual presentation. In this sense, the condition, in which a novel word was accompanied by PRG experienced reduction and thus a progressive abstraction of semantic features related to it. Our study also contributes novel findings to gestural research on language learning in children since the participants were older than previously studied (Rohlfing, 2019). Regarding nonverbal behavior and learning, it remains a question for further research whether reduction affects learning in other tasks similarly, for instance, in explicit learning situations such as math.

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 University of Paderborn. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article. Drawings: Copyright © 2013 Joy Katzmarzik leap4joy graphics; reprinted with permission.

AUTHOR CONTRIBUTIONS

UM and KR conceived and designed the study. UM piloted the study, recruited participants, conducted the study, and analyzed the data. UM and KR drafted the manuscript. Both authors commented on, edited, and revised the manuscript prior to submission. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the German Research Foundation (DFG) under the grant number to KR (RO 2443).

ACKNOWLEDGMENTS

We gratefully thank the schools, all children who participated in this study, and our research assistant Lina Kreutzkamp.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.651725/full#supplementary-material

Supplementary Figure 1 | The Figure shows the gesture versions and pictures of the target words "staksen", "retschen" and "krauchen" (Copyright © 2013 Joy Katzmarzik leap4joy graphics; reprinted with permission).

Supplementary Figure 2 | The Figure shows the gesture versions and pictures of the target words "gliddern", "Beisa" and "Ralle" (Copyright © 2013 Joy Katzmarzik leap4joy graphics; reprinted with permission).

Supplementary Figure 3 | The Figure shows the gesture versions and pictures of the target words "Alk" and "Fennek" (Copyright © 2013 Joy Katzmarzik leap4joy graphics; reprinted with permission).

REFERENCES

Arnold, J. E., Kaiser, E., Kahn, J., and Kim, L. (2013). Information structure: linguistic, cognitive, and processing approaches. *WIRES Cogn. Sci.* 4, 403–413. doi: 10.1002/wcs.1234

Aussems, S., and Kita, S. (2020). Seeing iconic gesture promotes first- and second-order verb generalization in preschoolers. *Child Dev.* 92, 124–141. doi: 10.1111/cdev.13392

Baddeley, A. (1986). *Working Memory: Oxford Psychology Series, no. 11*. Oxford: Oxford University Press.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *J. Mem. Lang.* 42, 1–22. doi: 10.1006/jmla.1999.2667

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang.* 60, 92–111. doi: 10.1016/j.jml.2008.06.003

Bohn, M., Kachel, G., and Tomasello, M. (2019). Young children spontaneously recreate core properties of language in a new modality. *Proc. Natl. Acad. Sci. U.S.A.* 116, 26072–26077. doi: 10.1073/pnas.1904871116

Capone, N. C., and McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *J. Speech. Lang. Hear. Res.* 48, 1468–1480. doi: 10.1044/1092-4388(2005/102)

Capone Singleton, N. C. (2012). Can semantic enrichment lead to naming in a word extension task? *Am. J. Speech Lang. Pathol.* 21, 279–292. doi: 10.1044/1058-0360(2012/11-0019)

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Lang. Cogn. Process.* 27, 594–610. doi: 10.1080/01690965.2011.567074

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.* 55, 51–86. doi: 10.1146/annurev.psych.55.090902.142050

Esteve-Gibert, N., and Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Commun.* 57, 301–316. doi: 10.1016/j.specom.2013.06.006

Forster, S., and Lavie, N. (2008). Failures to ignore entirely irrelevant distractors: the role of load. *J. Exp. Psychol.* 14, 73–83. doi: 10.1037/1076-898X.14.1.73

Fowler, C. A., Levy, E. T., and Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *J. Mem. Lang.* 37, 24–40. doi: 10.1006/jmla.1996.2504

Galati, A., and Brennan, S. E. (2010). Attenuating information in spoken communication: for the speaker, or for the addressee? *J. Mem. Lang.* 62, 35–51. doi: 10.1016/j.jml.2009.09.002

Galati, A., and Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Lang. Cogn. Neurosci.* 29, 435–451. doi: 10.1080/01690965.2013.796397

Gentner, D. (1982). "Why nouns are learned before verbs: linguistic relativity versus natural partitioning," in *Language Development: Language, Thought, and Culture*, ed. S. A. Kuczaj (Hillsdale, NJ: Lawrence Erlbaum), 301–334.

Gerwing, J., and Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture* 4, 157–195. doi: 10.1075/gest.4.2.04ger

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395

Goldstone, R. L., and Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *J. Learn. Sci.* 14, 69–110. doi: 10.1207/s15327809jls1401_4

Goodrich, W., and Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Dev. Sci.* 12, 81–87. doi: 10.1111/j.1467-7687.2008.00735.x

Gregory, M. L. (2002). *Linguistic Informativeness and Speech Production: An Investigation of Contextual and Discourse-Pragmatic Effects on Phonological Variation*. Ph.D. Dissertation. Boulder, CO: University of Colorado Boulder.

Griffin, Z. M., and Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *J. Mem. Lang.* 38, 313–338. doi: 10.1006/jmla.1997.2547

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cogn. Linguist.* 19, 1–33. doi: 10.1515/COG.2008.001

Heller, V., and Rohlfing, K. J. (2017). Reference as an interactive achievement: sequential and longitudinal analyses of labeling interactions in shared book reading and free play. *Front. Psychol.* 8:139. doi: 10.3389/fpsyg.2017.00139

Hirsh-Pasek, K., and Golinkoff, R. M. (2006). *Action Meets Word: How Children Learn Verbs*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780195170009.001.0001

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., and Swerts, M. (2015). Reduction in gesture during the production of repeated references. *J. Mem. Lang.* 7, 1–17. doi: 10.1016/j.jml.2014.10.004

Holler, J., and Wilkin, K. (2009). Communicating common ground: how mutually shared knowledge influences speech and gesture in a narrative task. *Lang. Cogn. Process.* 24, 267–289. doi: 10.1080/01690960802095545

Holler, J., and Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *J. Nonverbal Behav.* 35, 133–153. doi: 10.1007/s10919-011-0105-6

Horst, J. S., and Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy.* 13, 128–157. doi: 10.1080/15250000701795598

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495

Hostetter, A. B., and Alibali, M. W. (2019). Gesture as simulated action: revisiting the framework. *Psychon. Bull. Rev.* 26, 721–752. doi: 10.3758/s13423-018-1548-0

Jacobs, N., and Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *J. Mem. Lang.* 56, 291–303. doi: 10.1016/j.jml.2006.07.011

Jescheniak, J. D., and Levelt, W. J. M. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *J. Exp. Psychol.* 20, 824–843. doi: 10.1037/0278-7393.20.4.824

Jesse, A., and Johnson, E. K. (2012). Prosodic temporal alignment of co-speech gestures to speech facilitates referent resolution. *J. Exp. Psychol.* 38, 1567–1581. doi: 10.1037/a0027921

Kelly, S. D., Barr, D. J., Church, R. B., and Lynch, K. (1999). Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *J. Mem. Lang.* 40, 577–592. doi: 10.1006/jmla.1999.2634

Kelly, S. D., Creigh, P., and Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254

Kendon, A. (1972). "Some relationships between body motion and speech," in *Studies in Dyadic Communication*, eds A. W. Siegman and B. Pope (New York, NY: Pergamon Press), 177–210. doi: 10.1016/b978-0-08-015867-9.50013-7

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 (Hague: Mouton Publishers), 207–227. doi: 10.1515/9783110813098.207

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511807572

Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059

Koke, M. (2019). *"Aber Wie Macht Denn Eine Ente?" "Na, Wie Sieht Denn Eine Ente aus?" Gestenetablierung Und -Reduktion Bei 4-Jährigen Kindern im Kontext eines Gemeinsamen Ziels*. Ph.D. thesis. Paderborn: Paderborn University.

Lam, T. Q., and Watson, D. G. (2010). Repetition is easy: why repeated referents have reduced prominence. *Mem. Cogn.* 38, 1137–1146. doi: 10.3758/MC.38.8.1137

Lavelli, M., and Majorano, M. (2016). Spontaneous gesture production and lexical abilities in children with specific language impairment in a naming task. *J. Speech Lang. Hear. Res.* 59, 784–796. doi: 10.1044/2016_JSLHR-L-14-0356

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *J. Exp. Psychol.* 21, 451–468. doi: 10.1037/0096-1523.21.3.451

Lavie, N. (2005). Distracted and confused: selective attention under load. *Trends Cogn. Sci.* 9, 75–82. doi: 10.1016/j.tics.2004.12.004

McGregor, K. K., Friedman, R. M., Reilly, R. M., and Newman, R. M. (2002). Semantic representation and naming in young children. *J. Speech Lang. Hear. Res.* 45, 332–346. doi: 10.1044/1092-4388(2002/026)

McGregor, K. K., Rohlfing, K. J., Bean, A., and Marschner, E. (2009). Gesture as a support for word learning: the case of under. *J. Child Lang.* 36, 807–828. doi: 10.1017/S0305000908009173

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

McNeil, N. M., and Fyfe, E. R. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learn. Instr.* 22, 440–448. doi: 10.1016/j.learninstruc.2012.05.001

Mertens, U. (2017). *Erweiterung Des Wortschatzes von Vorschulkindern Unter der Verwendung sich Sukzessiv Reduzierender Ikonischer Gestik Mittels Eines Social Robots*. Ph.D. thesis. Bielefeld: Bielefeld University.

Mumford, K. H., and Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Dev.* 85, 1181–1189. doi: 10.1111/cdev.12188

Munro, N., Baker, E., McGregor, K., Docking, K., and Arciuli, J. (2012). Why word learning is not fast. *Front. Psychol.* 3:41. doi: 10.3389/fpsyg.2012.00041

Nachtigäller, K., Rohlfing, K. J., and McGregor, K. K. (2013). A story about a word: does narrative presentation promote learning of a spatial preposition in German two-year-olds? *J. Child Lang.* 40, 900–917. doi: 10.1017/S0305000912000311

Nomikou, I., Rohlfing, K. J., Cimiano, P., and Mandler, J. M. (2019). Evidence for early comprehension of action verbs. *Lang Learn Dev.* 15, 64–74. doi: 10.1080/15475441.2018.1520639

Nooijer, J. A., de van Gog, T., Paas, F., and Zwaan, R. A. (2014). Words in action: using gestures to improve verb learning in primary school children. *Gesture* 14, 46–69. doi: 10.1075/gest.14.1.03noo

Ping, R., and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cogn. Sci.* 34, 602–619. doi: 10.1111/j.1551-6709.2010.01102.x

Rees, G., Frith, C. D., and Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science* 278, 1616–1619. doi: 10.1126/science.278.5343.1616

Rohlfing, K. J. (2019). "Learning language from the use of gestures," in *International Handbook of Language Acquisition*, eds J. S. Horst and J. K. Torkildsen (New York, NY: Routledge), 213–233. doi: 10.4324/9781315110622-12

Son, J. Y., Smith, L. B., Goldstone, R. L., and Leslie, M. (2012). The importance of being interpreted: grounded words and children's relational reasoning. *Front. Psychol.* 3:45. doi: 10.3389/fpsyg.2012.00045

University of Leipzig (1998). *Wortschatzportal (Vocabulary Portal)*. Available online at: www.wortschatz.uni-leipzig.de (accessed July 6, 2017).

Vogt, P., de Haas, M., de Jong, C., Baxter, P., and Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Front. Hum. Neurosci.* 11:73. doi: 10.3389/fnhum.2017.00073

Vogt, S., and Kauschke, C. (2017a). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *J. Child Lang.* 44, 1458–1484. doi: 10.1017/S0305000916000647

Vogt, S., and Kauschke, C. (2017b). With some help from others' hands: iconic gesture helps semantic learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* 60, 3213–3225. doi: 10.1044/2017_JSLHR-L-17-0004

Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: an overview. *Speech Commun.* 57, 209–232. doi: 10.1016/j.specom.2013.09.008

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., and Goldin-Meadow, S. (2018a). Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci.* 21:e12664. doi: 10.1111/desc.12664

Wakefield, E. M., Hall, C., James, K. H., and Goldin-Meadow, S. (2018b). Gesture for generalization: gesture facilitates flexible learning of words for actions on objects. *Dev. Sci.* 21:e12656. doi: 10.1111/desc.12656

Watson, D. G., Arnold, J. E., and Tanenhaus, M. K. (2008). Tic Tac TOE: effects of predictability and importance on acoustic prominence in language production. *Cognition* 106, 1548–1557. doi: 10.1016/j.cognition.2007.06.009

Wojcik, E. H. (2013). Remembering new words: integrating early memory development into word learning. *Front. Psychol.* 4:151. doi: 10.3389/fpsyg.2013.00151

Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x

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

frontiers
in Psychology

REVIEW
published: 29 April 2021
doi: 10.3389/fpsyg.2021.634074

The Role of Iconic Gestures in Speech Comprehension: An Overview of Various Methodologies

Kendra G. Kandana Arachchige, Isabelle Simoes Loureiro, Wivine Blekic, Mandy Rossignol and Laurent Lefebvre*

Cognitive Psychology and Neuropsychology, University of Mons, Mons, Belgium

Iconic gesture-speech integration is a relatively recent field of investigation with numerous researchers studying its various aspects. The results obtained are just as diverse. The definition of iconic gestures is often overlooked in the interpretations of results. Furthermore, while most behavioral studies have demonstrated an advantage of bimodal presentation, brain activity studies show a diversity of results regarding the brain regions involved in the processing of this integration. Clinical studies also yield mixed results, some suggesting parallel processing channels, others a unique and integrated channel. This review aims to draw attention to the methodological variations in research on iconic gesture-speech integration and how they impact conclusions regarding the underlying phenomena. It will also attempt to draw together the findings from other relevant research and suggest potential areas for further investigation in order to better understand processes at play during speech integration process.

Keywords: iconic gestures, speech-gesture integration, methodological considerations, co-network connectivity, multisensory integration

OPEN ACCESS

Edited by:
Kazuki Sekine,
Waseda University, Japan

Reviewed by:
Seana Coulson,
University of California, San Diego,
United States
Nerea Aldunate Ruff,
Pontificia Universidad Católica
de Chile, Chile

****Correspondence:***
Kendra G. Kandana Arachchige
kendra.kandanaarachchige@
umons.ac.be

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

Received: *26 November 2020*
Accepted: *01 April 2021*
Published: *29 April 2021*

Citation:
Kandana Arachchige KG, Simoes Loureiro I, Blekic W, Rossignol M and Lefebvre L (2021) The Role of Iconic Gestures in Speech Comprehension: An Overview of Various Methodologies.
Front. Psychol. 12:634074.
doi: 10.3389/fpsyg.2021.634074

INTRODUCTION

"Gestures" refer to dynamic movements of the hands (Novack et al., 2016), with "iconic gestures" referring more precisely to manual movements allowing for the transmission of additional or redundant information to the speech they accompany (Kita and Özyürek, 2003; Willems et al., 2007). These gestures greatly contribute to the quality of the information exchange between individuals from an early age onwards. Since the 1990s, numerous attempts have been made to understand the mechanisms underlying the understanding of these gestures and their integration into the associated verbal utterance. Indeed, these gestures appear to possess semantic information that is related to the verbally conveyed message. The notion of "gesture-speech integration" is a central concept in this field. It refers to the implicit cognitive process of combining audio-visual information into a single representation (Green et al., 2009).

To date, studies on gesture-speech integration have employed diverse methodologies, whether in terms of the definition for iconic gestures used, the task or even instructions given to participants. And as suggested by Wolf et al. (2017), the interpretation of verbal and gestural information can be modulated according to the task and/or instruction given to the participants. Our aim is, therefore, to put into perspective the data found in the field of iconic gesture-speech integration by specifically highlighting the methodological variations. Indeed, the diversity of results yielded in this field could be explained by (1) non-identical testing methods (Wolf et al., 2017) (2) overlooking the specificities of iconic gestures or (3) a non-integrative interpretation of results.

First, an integrated and comprehensive definition of iconic gestures will be given to contextualize the focal point of this review. There will then be a focus on the different methodological variations when investigating the links between iconic and verbal information in behavioral, electrophysiological, brain imaging, and brain stimulation studies. Clinical population studies will also be discussed, as they can shed some light on the processes underlying the integration of gestural and verbal information. The discussion will then attempt to integrate all elements to suggest potential avenues for future studies and improve the understanding of the interrelation between iconic gestures and verbal language.

CHARACTERIZING ICONIC GESTURES: TOWARD AN INTEGRATED AND COMPREHENSIVE DEFINITION

Iconic gestures convey meaning semantically related to the content of the co-occurring speech (McNeill, 1992). This definition of iconic gestures can be found in the majority, if not all, of the studies conducted on gesture-speech integration. On its own, it might not be sufficient to describe the variety of iconic gestures. These gestures being the central focus of this review, it is proposed to focus on exactly what they represent. The literature identified an iconic gesture as:

(a) A meaningful manual movement (Kita and Özyürek, 2003; Willems et al., 2007);
(b) Temporally aligned to the speech it accompanies (McNeill, 1992; Willems et al., 2007; Habets et al., 2011; Obermeier and Gunter, 2014);
(c) Conveying redundant or complementary information to that present in the co-occurring speech (Krauss et al., 1996; Kita and Özyürek, 2003);
(d) Semi-automatically integrated with speech (Holle and Gunter, 2007; Kelly et al., 2010a);
(e) Providing information on actions (and is then called *kinetograph*), on the shape/size of an object (called *pictograph*), or on spatial relationship between two objects (called *spatial movement*);
(f) Carrying intrinsic meaning but rely on speech to be understood (Krauss et al., 1996; Hadar and Butterworth, 1997; Holle and Gunter, 2007);
(g) consisting of 3 phases (i.e., preparation-stroke-retraction), with the *stroke* carrying most of the semantic content (McNeill, 1992);

Given the variety of iconic gestures, it is essential to know exactly what is being investigated. This will be of a particular interest for this paper. Having these points in mind, the next section will focus on results obtained through various methodologies in behavioral, brain activity and brain stimulation investigations.

INVESTIGATING THE RELATIONSHIP BETWEEN ICONIC GESTURES AND LANGUAGE

Historically, two visions regarding the underlying processes involved in the comprehension and integration of iconic gestures with speech coexist in the literature. On the one hand, Krauss et al. (1991) considered iconic gestures as an epiphenomena of verbal language and do not consider them to have any relevant value in the understanding of the message. On the other hand, most studies and authors now argue in favor of the importance of iconic gestures in language comprehension (McNeill, 1992; Hadar and Butterworth, 1997; Beattie and Shovelton, 2002; Holler and Beattie, 2003; Kelly et al., 2010b), with some considering the gesture-speech integration to be automatic (McNeill, 1992; Kelly et al., 2004).

A recent meta-analysis (Dargue et al., 2019) investigated the effects of co-gesture on speech comprehension. Despite numerous studies showing an enhanced comprehension following the presentation of co-speech gestures, Dargue et al. (2019) highlighted only a moderate beneficial effect. The authors attributed this effect to the diverse methodologies used in the investigation of gesture-speech integration. However, they do not merely consider iconic gestures (the focus of this review) but also other types of co-speech gestures (such as deictic, metaphoric and beat gestures). Subsequently, the authors suggest to investigate the methodological variations within each type of co-speech gesture (Dargue et al., 2019). This review will, therefore, attempt to highlight these methodological aspects among iconic gestures.

First, iconic gesture-speech integration studies can be conducted through behavioral investigations, associated or not with a measure of brain activity or brain stimulation. Second, various experimental designs can be used to assess gesture-speech integration. One way is to modulate the relationship between the iconic gesture and the co-occurring speech. Three types of relationships may be of interest; (a) The information conveyed through iconic gestures may be *redundant* to that conveyed in speech, thereby reinforcing the message. For example, when speaking of a large object, the arm and hand gesture at an increasingly larger amplitude representing the width of the object. (b) Iconic gestures can also be *complementary* and thereby provide additional information to that contained in speech. For example, when speaking of a box one can gesture its shape. (c) The iconic gesture can also contradict the information contained in speech (Dick et al., 2014). In this case, the literature refers to an *incongruency*, most often semantic, between the verbal and gestural information [e.g., gesturing *stirring* while saying *break* (Willems et al., 2007)]. Manipulating the degree of congruency allows to take into account the semantical integration of information present in both modalities (Holle and Gunter, 2007). According to Holle and Gunter (2007), a decrease in performance following the presentation of incongruent information (represented by more incorrect responses or longer reaction times) can be interpreted as a failed attempt to integrate the gestural and verbal information.

Third, the task in itself can modulate the interpretation of results. Some studies require participants to simply observe the stimuli, whereas others require an explicit processing of the information by either focusing their attention on the verbal or gestural information. While observing the stimuli is an ecologically valid approach, focusing on one or the other aspect of the stimuli could seem less natural.

Fourth, investigating different types of iconic gestures could yield different results. As has been mentioned above, iconic gestures can represent actions, manner of movement or physical attributes (McNeill, 1992). More recently, Dargue and Sweller (2018b) also distinguished between typical and atypical iconic gestures.

Finally, other parameters can also be manipulated, such as the type of stimuli presented (i.e., recorded video clips of people gesturing, cartoons, or live presentation of gestures), their content (i.e., presenting single words, sentences, or a narrative), the length of the presented gesture (i.e., the complete gesture or just the stroke), or the visibility of the actor (i.e., if the face is made visible or masked).

The following section will review the literature considering these different parameters.

Behavioral Investigation of Gesture-Speech Integration

One way to investigate iconic gesture-speech integration is by varying the relationship between the iconic gesture and the co-occurring speech. To the best of the authors' knowledge, in the gestural domain, Church and Goldin-Meadow (1986) were the first to investigate discrepancies found between produced gestural movements and spoken words. Since then, many authors have contrasted the presentation of congruent vs. incongruent information to investigate the degree of integration between gestural and verbal information (Cassell et al., 1999; Kelly et al., 2004, 2010a; Wu and Coulson, 2005; Wu and Coulson, 2007a,b; Margiotoudi et al., 2014).

All behavioral studies manipulating gesture-speech congruency have highlighted faster reaction times and more correct responses when participants were in presence of congruent pairs compared to incongruent pairs (Kelly et al., 2010a,b; Margiotoudi et al., 2014; Wu and Coulson, 2014; Kandana Arachchige et al., 2018; Zhao et al., 2018; Momsen et al., 2020). These results were found when there was no specific task required (Green et al., 2009; Drijvers and Özyürek, 2018), as well as when participants were required to perform a task where they had to pay attention to the gesture (Kelly et al., 2010b; Margiotoudi et al., 2014; Cohen-Maximov et al., 2015; Nagels et al., 2019; Bohn et al., 2020; Özer and Göksun, 2020b), the speech (Ping et al., 2014; Wu and Coulson, 2014; Drijvers and Özyürek, 2018) or an un-related aspect (Kelly et al., 2010a; Kandana Arachchige et al., 2018; Zhao et al., 2018). Interestingly, a study using a priming paradigm failed to observe a congruency advantage on reaction times when participants were asked to match a target word to a gesture video prime (Wu and Coulson, 2007b). Here, the gesture primes were devoid of any accompanying speech. Since iconic gestures co-occur with a verbal utterance, an essential characteristic is missing for the gestures to be fully considered as *iconic*.

Seeing that the presence of an incongruent iconic gesture appears to hinder performance, whether it is attended to or not, Kelly et al. (2010b) suggested the presence of an obligatory and automatic integration between the two pieces of information. Since then, this automaticity has been put into perspective following data obtained through brain activity investigation. This will be discussed in the following section.

While the investigation of semantic (in)congruency constitutes a big part of the literature, numerous studies have contrasted the unimodal presentation of information (i.e., presenting gesture or speech alone) with a congruent bimodal presentation (i.e., presenting congruent information through both the gestural and verbal modalities) (Beattie and Shovelton, 2001; So et al., 2013; Iani and Bucciarelli, 2017). In a free-recall task, Beattie and Shovelton (2001) and Iani and Bucciarelli (2017) showed an increased information uptake following the presentation of bimodal compared to unimodal information. Yet, using a priming paradigm along with a lexical decision task, So et al. (2013) found no such advantage. It follows that three possible explanations for these contradictory results can be proposed. First, in the latter, it appears that the presented video clips were soundless. As mentioned above, an iconic gesture occurs concurrently to speech. The absence of speech during the video presentation could explain the lack of extra information. Second, participants were asked to respond to a written target. Although the neural correlates involved in the comprehension of spoken and visually presented words appear to overlap (Price et al., 1999), the temporality of the processing involved diverges (Marslen-Wilson, 1984). Third, the type of iconic gestures used in these studies differs. Beattie and Shovelton (2001) and Holler et al. (2009) showed that iconic gestures depicting physical attributes such as relative position, size or shape conveyed more information than other types. More recently, Dargue and Sweller (2018b) distinguished between typical and atypical iconic gestures, the former appearing to be more beneficial to speech comprehension.

The advantage of a congruent bimodal presentation is most noticeable with children. To understand when the ability of integrating gestural with verbal information develops, studies have investigated gesture-speech integration among children. Studies show that by the age of 3, children are capable of integrating iconic gestures representing physical attributes of objects with speech (Stanfield et al., 2013; Macoun and Sweller, 2016; Dargue and Sweller, 2018a; Aussems and Kita, 2019). The ability to integrate action iconic gestures appears to depend on the type of stimuli presentation used. When presenting video clips, Glasser et al. (2018) observed that children from the age of 4 were able to integrate the information from an action iconic gesture with speech to select a corresponding animated clip. Sekine et al. (2015) showed that children from the age of 3 were able to do so when the gestures were presented face-to-face. This real-life presentation advantage has also been observed for adults, particularly for iconic gestures depicting size and position (Holler et al., 2009).

Furthermore, by the age of 5, children presented with live action iconic gestures are able to recall more information compared to when presented with meaningless or no gestures (Kartalkanat and Göksun, 2020). These results were not shown for 3 year olds (Sekine et al., 2015). The age difference between these two studies could here be explained by the nature of the task, children having to pick a picture in the former study (Sekine et al., 2015) and having to produce an explicit answer in the latter (Kartalkanat and Göksun, 2020). In addition, research has shown that from the age of 3, children are able to understand the meaning behind an action iconic gesture in order to open a box in front of them (Bohn et al., 2020). This result was found whether the gesture was presented live or through a video clip. Miyake and Sugimura (2018) observed that the use of directive words (i.e., words indicating in which way an action is carried out) allowed for a better integration of information for 4 year olds. However, the absence of a "Gesture + Speech in the absence of directive words" condition makes it difficult to draw a definitive conclusion. Finally, among iconic gestures, just as for adults (Dargue and Sweller, 2018b), Dargue and Sweller (2020) highlighted that typical iconic gestures benefited comprehension compared to atypical iconic gestures for children.

In contrast to the development of the ability to integrate iconic gesture with speech in children, older adults appear to rely less on gestural information (Cocks et al., 2011). Developing on the suggestion by Thompson and Guzman (Thompson and Guzman, 1999), Cocks et al. (2011) suggested that a weakening of working memory capacities found in aging could explain the difficulty to focus on two different sources of information. More recently, Schubotz et al. (2019) found that older participants, unlike younger ones, did not adapt their words or gestures in a context of shared experience and conveyed less multimodal information when communicating. The results of these studies thus suggest an impairment in the ability to integrate iconic gestures together with speech, which could mirror the capacities developed during childhood (Cocks et al., 2011).

Another population that seems to benefit from a bimodal presentation of information is non-native speakers (Dahl and Ludvigsen, 2014). Dahl and Ludvigsen (2014) and Drijvers et al. (2019) observed an improved understanding of scene descriptions for non-native speakers when they were presented with action iconic gestures depicting physical attributes. By evaluating long-term information retrieval, Kelly et al. (2009) demonstrated that when participants needed to recall words in a foreign language, performances were facilitated when they were exposed to action iconic gestures during the encoding phase (e.g., they found that learning the word *drink* is easier when accompanied by the gesture representing the act of drinking). Other authors have also demonstrated that when presented with degraded verbal information, action iconic gestures improved the comprehension of verbs for non-natives speakers (Drijvers and Özyürek, 2020).

Finally, regarding population, one aspect that has recently started to be taken into account is individual differences. In a recent review, Özer and Göksun (2020a) plead for an assessment of individual differences in the field of gesture comprehension. Indeed, individuals vary regarding their verbal and visual-spatial abilities (Alfred and Kraemer, 2017) and iconic gestures appear to rely on these to be processed (Wu and Coulson, 2014). Given the on-line nature of gesture-speech integration, Wu and Coulson (2014) sought to investigate the involvement of working memory in gesture-speech integration. Using a dual-task paradigm, they showed that visual-spatial, but not verbal, working memory was involved in gesture-speech integration with a higher load on visual-spatial working memory affecting performances on the gesture-speech integration task (Wu and Coulson, 2014; Momsen et al., 2020). An iconic gesture containing semantically related information (McNeill, 1992), the absence of verbal working memory involvement is curious. One potential explanation would consist of not having considered individual differences when loading the verbal working memory span. In fact, the verbal high load condition on the secondary task was completed by having participants remembering 4 numbers (Wu and Coulson, 2014; Momsen et al., 2020). This was the same across all participants, whilst working memory abilities vary across individuals (Jarrold and Towse, 2006). Further research in this field could assess individual differences in a preliminary task and select an appropriate secondary task.

Overall, behavioral studies have highlighted (1) an advantage of congruent bimodal compared to unimodal presentation of information and (2) that iconic gestures seem to be processed in a parallel and automatic fashion with the speech it accompanies. While light variations in individual results can be found, these can be explained by variations in methodological aspects or by not taking individual differences into account.

Beyond the afore-mentioned studies, another large part of the literature has aimed to understand gesture-speech integration within the framework of imaging, electrophysiology and brain stimulation research.

Investigating Brain Activity During Gesture-Speech Integration

While behavioral studies highlight the interest of adding iconic gestures to speech to enhance observable and quantifiable performance, brain activity can help determine when and where this integration of information takes place. In fact, research in this area is vast. Electrophysiological studies can help reveal the temporal aspects of gesture-speech integration while brain imaging and stimulation studies can shed light on where the integration is taking place. Additionally, just as in behavioral investigations, studies can manipulate the relationship between speech and gesture, use different types of iconic gestures, investigate different populations, etc.

This section will first review electrophysiological studies before focusing on brain imaging and brain stimulation studies.

Electrophysiological Studies

As mentioned previously, these studies allow for a temporal approach to semantic integration. More specifically, event-related potentials provide information on the temporal course of the neuronal processes involved following the presentation of a sensory stimulus (Srinivasan, 2005).

Whilst behavioral studies contrasting a congruent and incongruent presentation of information suggested the presence of an automatic integration, electrophysiological studies have highlighted different brain responses depending on whether congruent or incongruent information was presented (Özyürek et al., 2007; Kelly et al., 2010a; Habets et al., 2011).

Studies by Özyürek et al. (2007), Kelly et al. (2010a), and Habets et al. (2011) have demonstrated a larger N400 component following the presentation of incongruent compared to congruent iconic gesture-speech pairs. These studies all investigated action iconic gestures. They did not require the participants to direct their attention to either speech or gesture and presented video clips of the stroke without making the actor's face visible. According to Holcomb (1993), the N400 component allows to measure the effort required to unify each presented item into an integrated representation. An increase in the N400 component amplitude would, therefore, appear as a complication of this process. Holcomb further suggests that the N400 component reflects a process between the recognition and integration processes (i.e., an activation in a post-semantic memory system). Other authors have suggested that it can reveal a semantic violation in a given context (Luck, 2014), index the level of difficulty to retrieve the associated conceptual representation (Kutas et al., 2006) and arise from a series of processes activating and integrating the target item's meaning into the presented context (Nieuwland et al., 2020). This component is generally observed in language studies (Kutas and Federmeier, 2011) but can also be elicited by non-linguistic stimuli (Sitnikova et al., 2003).

The presence of a larger N400 component for the incongruent pairs in gesture-speech integration studies (Özyürek et al., 2007; Kelly et al., 2010a; Habets et al., 2011) could thus suggest a difficulty in the semantic processing for these pairs and/or a difficulty in integrating the activated meanings into one unified representation. However, while two studies investigated the incongruency on single words (Habets et al., 2011; Kelly et al., 2010a), the third investigated the incongruency effect within a sentence context (Özyürek et al., 2007). This methodological variation could account for the distinct N400 component site in the three studies. The two studies focusing on single words elicited the largest N400 component in the centro-parietal region, whereas the third study found the largest amplitude in more anterior regions. Using a dual task paradigm, Momsen et al. (2020)'s study also showed the presence of a N400 component being at its largest over anterior channels when presenting sentences. The anterior location is compatible with previous language research eliciting a larger N400 component over anterior regions when in presence of a semantic violation in a sentence context (Hald et al., 2006). This explanation is consistent with the results from another study contrasting ERPs elicited by speech accompanied or not by iconic gestures in a sentence context (Wu and Coulson, 2010). This study showed a larger N400 component over central and centroparietal regions in the absence of iconic gestures (Wu and Coulson, 2010). While the centroparietal effect was found in a sentence context, it was elicited by the absence of an iconic gesture, rather than an incongruent iconic gesture (Momsen et al., 2020). The centroparietal regions, therefore, appear to be involved at a local integration level while anterior regions appear to deal with a global sentence-level integration.

These results led Bernardis et al. (2008) to suggest that the presence of an incongruency slows down the activation of meanings. In line with Thompson and Guzman (1999) and Cocks et al. (2011) proposed that when the presented information was incongruent, the meanings could not be integrated into the working memory, consequently modifying brain activity (Bernardis et al., 2008).

An increase of the N400 component has also been observed when an incongruency was present between a soundless gesture clip and an unrelated word, even when the latter occurred one second after the offset of the gesture clip (Wu and Coulson, 2007b). This result, along with others that highlight the presence of an increased N400 component, despite a long inter-stimulus-interval (Kelly et al., 2004; Wu and Coulson, 2005), appear contradictory to the study by Habets et al. (2011). This research demonstrated that when a gesture and its corresponding utterance were presented 360 ms apart, the incongruency effect reflected by an increased N400 component was not present (Habets et al., 2011). One potential explanation resides in the nature of the stimuli. Wu and Coulson (2007b) presented stimuli that could have been less ambiguous given that in a previous task participants were required to explicitly judge their relatedness to gestures. The stimuli used in Habets et al. (2011)'s study were more ambiguous and hardly understandable without speech.

This incongruency effect was also found in subsequent studies, eliciting a N450 component. This component is thought to be equivalent to the N400 component but specific to a visual/gestural stimulus (Wu and Coulson, 2005, 2007a,b). Just as the N400, it seems to be influenced by the degree of congruency between the iconic gestures and the context in which they are presented (Wu and Coulson, 2005). Indeed, Wu and Coulson (2005) observed an increase of the N450 amplitude when iconic gesture videos (representing either actions or physical attributes of objects) were incongruent to previously presented cartoons. This result was then replicated in the same study when participants were required to relate a target word to the previously presented context (Wu and Coulson, 2005). And in another study assessing the congruency effect between a prime iconic gesture video and target word (Wu and Coulson, 2007b). Interestingly, the N450 component has essentially been demonstrated in studies where the gestures were presented as soundless video clips. This is compatible with the vision of the N450 component as specific to visual stimuli (Wu and Coulson, 2005), as in the absence of speech, the gesture video becomes a visual stimulus.

Furthermore, Holle and Gunter (2007) observed a larger N400 component when an iconic gesture supporting the high frequency homonym or a meaningless gesture followed a low frequency verbal homonym. According to the authors, this suggests that the iconic gesture was able to facilitate the processing of the low frequency homonym. Therefore, by varying the type of gesture presented, Holle and Gunter (2007) demonstrated that iconic gestures can facilitate speech comprehension when the latter needs to be disambiguated. In addition, they questioned the automaticity of gesture-speech integration following the disturbance caused by meaningless grooming gestures (Holle and Gunter, 2007). As attested by these authors, should the integration really be automatic, the presence of grooming movements should not have modified performances. Consistent with this, Kelly et al. (2007) demonstrated that the N400 component to incongruent stimuli could also be modulated by the presence of knowledge on the intentional relationship between gesture and speech. In this case, they found a larger amplitude of the N400 component for incongruent stimuli when participants were aware of the mismatch between the actor uttering the sentence and the one performing the gesture.

Another discrepancy found in the literature concerns early effects. Kelly et al. (2004) reported early sensory effects through a fluctuation of the P1, N1, and P2 components. The P1 component is modulated by selective attention and state of alertness (Luck et al., 2000), the N1 component is influenced by spatial aspects of the stimulus (Mangun, 1995; Hillyard and Anllo-Vento, 1998), and the P2 reflects perceptual processing (Luck and Kappenman, 2011). Kelly et al. (2004) interpreted the presence of these early effects as the creation, through gestures, of a visual-spatial context affecting language processing. According to the authors, the visibility of the actor's face could have allowed these effects. However, no early effects were found in other studies presenting a visible actor's face (Wu and Coulson, 2005, 2007a,b). Another explanation could reside in the complexity of the stimuli. In their study, Kelly et al. (2004) repeatedly used the same four simple stimuli (i.e., tall, thin, short, and wide). This repetition could have favored the creation of an expected visual context, thereby eliciting early effects.

Finally, electrophysiological studies have also been conducted on non-native speakers and, recently, children. For non-native speakers, Drijvers and Özyürek (2018) observed a larger N400 component for incongruent stimuli pairs. This effect disappeared in the event of degraded speech for non-natives, but remained for native speakers (Drijvers and Özyürek, 2018). The authors theorized that a minimum quality of the auditory stimulus is required for the integration process to take place for non-native listeners (Drijvers and Özyürek, 2018). A subsequent study corroborated these results, revealing that unlike for native speakers, non-native speakers do not benefit from visible speech (i.e., visible phonological information) in a degraded auditory context (Drijvers and Özyürek, 2020).

To the best of the authors' knowledge, only one electrophysiological study investigating gesture-speech integration in children has been conducted. In their study, Sekine et al. (2020) observed a larger N400 component for the incongruent trials compared to congruent ones. In line with data from behavioral studies on the development of gesture-speech integration in children (Stanfield et al., 2013; Sekine et al., 2015; Glasser et al., 2018), this study suggests that by the age of 6, children possess a qualitatively similar processing of gesture-speech information to adults.

In conclusion, although a late semantic effect has consistently been elicited, the same cannot be said for early effects. Other than for non-native speakers, results plead in favor of the existence of a semantic link between the iconic gestures and co-occurring verbal utterance. Electrophysiological studies thus corroborate results from behavioral studies. The absence of consistent results relating to early effects could be explained by the type of iconic gesture presented. As highlighted, iconic gestures comprise a variety of more or less complex gestural movements and can be redundant or complementary to speech. Moreover, the presence of late semantic effects is not exclusive to the presentation of iconic gestures. Consequently, although this constitutes a good first step in understanding the neural process involved in iconic gesture-speech integration, further investigation is required to deepen an understanding of this research area.

Brain Imaging and Brain Stimulation Studies

One way to enhance our understanding of the neural processes involved in gesture-speech integration is by using functional Magnetic Resonance Imaging (fMRI). This would allow to highlight which brain regions are involved in the processing of iconic gestures and understand their relationship to speech.

Most of the fMRI studies have been conducted with simple observation tasks (Willems et al., 2007; Dick et al., 2009, 2012, 2014; Green et al., 2009; Holle et al., 2010; Straube et al., 2011; Demir-Lira et al., 2018). Wolf et al. (2017) justified this choice by highlighting the possible motor-related artifacts caused by participants having to produce a motor response. In fact, Willems et al. (2007) observed an involvement of typical motor areas (such as the premotor cortex) in language processing, and typical language areas (such as Broca's area) in action processing. A motor involvement has also been suggested by behavioral studies (Ping et al., 2014; Iani and Bucciarelli, 2017). These studies showed that hand/arm movements produced by the participants hindered their ability to integrate gesture-speech information. Interestingly, this interference effect was not observed when participants were required to move their foot/leg (Ping et al., 2014; Iani and Bucciarelli, 2017).

With regard to gesture-speech integration, three main regions were found to be involved: the left inferior frontal gyrus (left IFG), the middle temporal gyrus (MTG) and the posterior superior temporal sulcus (pSTS).

An increase in the activity of the left IFG was observed during the presentation of iconic gestures (Dick et al., 2009) when they were incongruent to speech (Willems et al., 2007, 2009) and when they conveyed complementarity (Holler et al., 2015) compared to redundant information (Dick et al., 2014). But this enhanced activity was not found when comparing the presence and absence of iconic gestures (i.e., comparing a Gesture + Speech condition to a Gesture + Unrelated Movement or to a Speech Alone condition) (Holle et al., 2008; Dick et al., 2009; Green et al., 2009; Straube et al., 2011). The involvement of the left IFG in gesture-speech integration, therefore, appears to not merely be restricted to combining information, but rather to detect incompatibilities (Willems et al., 2007, 2009) and/or create a new coherent representation from two ambiguous inputs (Dick et al., 2014; Holler et al., 2015). This is consistent with viewing the left IFG as a unification site (Zhu et al., 2012). The process of unification allows for either lexically retrieved information, or meanings extracted from non-linguistic modalities to be integrated into one representation (Hagoort et al., 2009). Studies in the language domain suggest a functional separation between anterior and posterior regions of the IFG, the former being linked to controlled semantic retrieval and the latter to general selection processes (Gough et al., 2005; Humphries et al., 2007; Lau et al., 2008; Hagoort et al., 2009). Consequently, one possibility would be to allocate the integration of complementary information to anterior regions and the processing of incongruency to the posterior regions.

The role of the IFG as a unification rather than an integration site could explain why its activation is not limited to iconic gestures. Indeed, Willems et al. (2009) highlighted an increase of the left IFG activation following the presentation of incongruent pantomimes. Straube et al. (2011) observed an increased activation for congruent metaphorical but not iconic gestures. The explanation would here reside in the higher effort needed to comprehend metaphorical gestures as they represent abstract concepts.

Therefore, rather than being exclusive to iconic gesture processing, the left IFG is involved when (1) a deeper processing of information is required, (2) and a new representation of the information must be created and/or (3) when there is an incompatibility between several representations that needs to be resolved. In a recent study investigating the effects of transcranial magnetic stimulation (TMS), Zhao et al. (2018) caused slower reaction times on a gesture-speech integration task after stimulating (and therefore disrupting) the left IFG and left pMTG. TMS is a non-invasive neuro-stimulation technique that disrupts neuronal activity by inducing a virtual lesion (Pascual-Leone et al., 2000). This highlights the cortical areas involved in a task and the temporality at which this contribution takes place (Hallett, 2000) and demonstrates a causal relationship between a neural process and the behavior observed on the task (Rossini and Rossi, 2007). Hence, results from Zhao et al. (2018) suggest a reduction in the integration of iconic gestural and verbal information following a disruption of the left IFG and pMTG. Because these effects were obtained through two different protocols, the authors suggested that the IFG and pMTG contribute to gesture-speech integration, respectively, to retrieve contextual semantic information and stored semantic information (Zhao et al., 2018).

An involvement of the MTG in gesture processing has also been put forward by several studies (Dick et al., 2009, 2014; Willems et al., 2009; Holler et al., 2015; Demir-Lira et al., 2018). Still, just as for the IFG, results vary depending on the nature of the stimuli. Dick et al. (2009) observed an increase in bilateral MTG activity in the presence of gestures though it did not discriminate between co-speech gestures and meaningless gestures. A specific activity increase was highlighted by Willems et al. (2009) for incongruent pantomimes (i.e., gestures that can be understood in the absence of any speech) but not for incongruent iconic gestures. However, when investigating the effects of complementary iconic gestures (rather than redundant), several studies have demonstrated an increased MTG activation (Dick et al., 2014; Holler et al., 2015; Demir-Lira et al., 2018). While Dick et al. (2014) highlighted this increased activity on the left MTG for adults, Demir-Lira et al. (2018) observed it on the right MTG for children. The difference of location has been suggested to reflect the possible use of additional cues in children compared to adults (Demir-Lira et al., 2018). Finally, Holler et al. (2015) observed that when listeners were specifically addressed, the presence of iconic complementary gestures elicited an increased right MTG activation.

Wagner et al. (2001) suggested that the left MTG could work together with the left IFG to retrieve semantic information (Kuperberg et al., 2008). Although the IFG appears to be sensitive to congruency (Willems et al., 2009), the MTG does not. In the language domain, Badre et al. (2005) suggested that the MTG was sensitive to target association but not competition. This is consistent with an involvement of the MTG in integrating complementary iconic information with speech.

The third main site that appears to be involved in gesture-speech integration is the left pSTS (Holle et al., 2008; Straube et al., 2011; Demir-Lira et al., 2018). In the field of recognition, this region appears to be involved in the integration of multimodal information (Beauchamp et al., 2004a,b). In language comprehension, the STS is activated during speech presentation (Crinion et al., 2003) with the left temporal cortex critically involved in the storage and retrieval of linguistic information (Hagoort, 2013). In the gesture-speech integration domain, studies have again yielded mixed results. In some studies, although an increased activation of the left STS was found in the "Speech + Gestures" condition, this activation either wasn't sensitive to the meaning of gestures (Willems et al., 2007; Dick et al., 2009, 2012, 2014), or was greater in the case of incongruent pantomimes but not iconic gestures (Willems et al., 2009). These latter results, along with the observed activation of MTG for pantomimes, led Willems et al. (2009) to suggest that pSTS/MTG was involved in the integration of information on a relatively stable conceptual representation. According to the authors, the nature of co-speech gestures (i.e., language-dependent) require that they be integrated at a higher level, given that they involve the creation of a novel representation.

Interestingly, this very explanation was later taken up by Straube et al. (2011) to explain the presence of a greater left pSTG (posterior superior temporal gyrus) activity in the "Speech + Iconic" and "Speech + Metaphoric gestures" conditions compared to Speech Alone. Though these authors offer the same role of pSTS/pSTG, they seem to disagree on which co-speech gesture it processes. Other studies have shown an involvement of the left pSTS in iconic gesture processing. Comparing the presence of iconic gestures to grooming movements, Holle et al. (2008) highlighted a greater activation of the left pSTS for the former. A different study replicated these findings by observing a bimodal enhancement over the pSTS/STG region when in presence of "Speech + Iconic gestures" (Holle et al., 2010). It also observed that this augmentation was greater in the context of degraded speech (Holle et al., 2010). Similarly, a previous study showed an increased activation of left superior temporal areas when the presented speech mismatched the sentence context (Willems et al., 2007). Holle et al. (2010) purported the existence of a sensitivity gradient within the pSTS/STG. This, with anterior portions being sensitive to speech processing and posterior regions (near the temporo-occipital, TO, junction) being sensitive to gestural information (Holle et al., 2010). This is consistent with a study by Green et al. (2009) that demonstrated an augmented activation at the left TO junction in the presence of Familiar Speech + Iconic gestures.

Brain imaging studies investigating gesture-speech integration in children are rare. When comparing the presence of iconic gestures, metaphoric gestures and grooming movements, Dick et al. (2012) observed an enhanced left pSTS activation for all types of movements relative to a baseline fixation activity. More recently, Demir-Lira et al. (2018) highlighted an increased left pSTS activity for complementary iconic gestures compared to redundant or no gestures. Because Dick et al. (2012) have not detailed the type of iconic gesture used or the relationship between the iconic gestures and speech (i.e., whether they were redundant or complementary), a direct and definitive comparison would be speculative. Furthermore, it is possible that limiting their sample to 9 children did not allow to investigate precise activation differences. Another possible explanation resides in the presence of methodological dissimilarities (Holle et al., 2008). More precisely, as the authors have highlighted, the relationship between gesture and speech as well as their level of integration could be key. It is possible that the pSTS serves as a local integration site [i.e., when the gesture is required to be integrated with the verbal unit (Holle et al., 2008)], and the IFG would act as a global integration site (i.e., where integration is required on a sentence level) (Willems et al., 2007; Dick et al., 2009; Holle et al., 2008, 2010). This supports the presence of a pSTS activation for complementary iconic gestures, the integration taking place on a local unit level.

Two main conclusions can be drawn from these findings. First, given the methodological variations (such as tasks, type of gesture or relationship between gesture and speech), defining one precise neural network involved in iconic gestures/speech comprehension is laborious. Yet, this variation can be beneficial for a more precise understanding of what is involved when, during iconic gesture-speech integration. Second, because these three areas (i.e., IFG, pSTS, and MTG) appear to be involved in various degrees and at different moments, connectivity studies could shed some light on the matter.

Hein and Knight (2008) suggested that the function of STS varies according to the nature of the co-activated network. This vision supports the idea that the same brain region can result in different cognitive processes depending on the nature of the task or stimuli involved. The existence of a task-dependent co-activated network reconciles the numerous observations mentioned hereinabove.

Recent studies have investigated the connectivity signature of co-speech gesture integration (Straube et al., 2018) and the spatial-temporal dynamics of gesture-speech integration (He et al., 2018). While their results support the key role of pSTS (He et al., 2018; Straube et al., 2018) and IFG (He et al., 2018) in gesture-speech integration, the gestures they investigated "could be comprehended even without accompanying speech" (Straube et al., 2018). Therefore, this does not fit the criteria to be classified as *iconic* gestures. Future research could attempt to explore the connectivity signature of iconic gestures integration.

Similarly, Drijvers et al. (2018) investigated the spatiotemporal changes in cerebral oscillations when the presence of gestures enhances clear or degraded speech. The study of brain oscillations has regained interest in the last decade (Wang, 2003; Ward, 2003; Weiss and Mueller, 2012; Başar, 2013) as it can provide complementary data to those obtained via fMRI on how brain activity relates to cognitive performances (Ward, 2003). A suppression of alpha and beta activity is found in regions that are engaged in a task (Jensen and Mazaheri, 2010; Quandt et al., 2012), while an increase in gamma activity is linked to an enhanced cognitive activity (Fitzgibbon et al., 2004; Jensen et al., 2007). Previous research has shown differentiated alpha and beta rhythms whether the gesture observed was iconic or deictic. This is consistent with alpha and beta rhythms being closely linked to the allocation of visual-spatial attention (Quandt et al., 2012) and that iconic and deictic gestures are processed differently. When gestures enhanced communication in a degraded speech context, Drijvers et al. (2018) demonstrated a greater suppression of alpha and beta activity over motor regions (hand motor area and supplementary motor area). According to the authors, this could suggest an attempt of imagining the action to aid comprehension (Drijvers et al., 2018). This is in agreement with a previous study showing alpha and beta power suppression in the precentral gyrus regions during motor imagery (De Lange et al., 2008). An alpha and beta suppression in frontal regions (Momsen et al., 2021) and more specifically in the left IFG and left pSTS/MTG, STG regions (Drijvers et al., 2018) is consistent with their role in gesture-speech integration highlighted by imaging studies. An increase in gamma power in the left temporal lobe was found at the presentation of the gesture's stroke and co-occurring speech, suggesting an attempt to integrate both information (Drijvers et al., 2018).

Overall, results in brain activity studies show the importance of knowing exactly what type of gesture is involved and its relationship to language. We have underlined that these two variables, along with the task involved, can modulate the interpretation given to the results and could explain apparent discrepancies between studies. In electrophysiological studies, the complexity of the presented task and iconic gesture can influence whether or not early sensory components are modulated. Mismatch paradigms consistently elicited the presence of a late semantic component. Yet, this component varies in its timing (N400–N450). We suggest that this variation was due to the stimuli that were used in the tasks (e.g., soundless video clips, audio-visual gestures). Brain imaging studies variously showed an involvement of the left IFG, left pSTS and MTG in gesture-speech integration. These activations appear to mainly depend on the nature of the relationship between iconic gesture and speech (i.e., redundant, complementary, or incongruent), as well as on the task. These variations plead in favor of the existence of a task-dependent co-activated network.

Investigating Gesture-Speech Integration in Clinical Populations

The study of behavior and cognition of clinical population allows for a better understanding of healthy cognition (Eysenck, 2014). The presence of an impairment of gesture-speech integration in patients could thus improve the understanding of the processes underlying the same integration in neurologically intact individuals. However, seemingly inconsistent results have also been highlighted within the same clinical population. This section will attempt to reconcile these apparent discrepancies by focusing on four clinical groups: aphasia, specific language impairment, autism spectrum disorder and schizophrenia.

Aphasia is an acquired disorder that can affect both language production and comprehension (Preisig et al., 2018). While language and gesture production have been vastly studied, literature on gesture comprehension is quite sparse.

In 1972, Gainotti and Ibba observed an impairment of pantomime comprehension among aphasic patients. This result was later replicated for aphasic patients presenting mono-hemispheric cerebral lesions compared to healthy participants and non-aphasic brain damaged patients (Gainotti and Lemmo, 1976). While these were among the first studies to focus on gesture comprehension in aphasia, they do not investigate iconic gestures, nor specify the type of aphasia involved. A couple of years later, a new study investigating pantomime processing showed that performances depended on the type of aphasia (Ferro et al., 1980). Ferro et al. (1980) showed that patients with Global, Wernicke and Transcortical aphasia presented lower performances at the Gesture Recognition task, compared to patients with Broca, Anomic or Conduction aphasia. The authors associated these results with the presence of lesions in the left posterior regions, involved in gesture identification. Since then, to the best of the authors' knowledge, studies have not differentiated their results according to aphasia type.

More recently, several studies investigated co-speech gesture-speech integration in aphasia (Eggenberger et al., 2016; Cocks et al., 2018; Preisig et al., 2018). Preisig et al. (2018) showed that during live conversations, co-speech gestures (of all types) attracted the attention of aphasic patients and were more fixated than abstract gestures. The authors suggested that these patients may benefit from the bimodal information presentation to compensate a verbal deficit (Preisig et al., 2018). However, because no task was involved, it is unclear whether patients understood and processed the meaning of these gestures. Another study required patients to explicitly integrate the iconic gesture meaning with the co-occurring speech by deciding whether they were congruent or not (Eggenberger et al., 2016). Results showed that patients performed better when presented with congruent compared to incongruent pairs or associated with meaningless movements. Eggenberger et al. (2016), therefore, concluded that congruent iconic gestures could enhance comprehension for patients with aphasia. Cocks et al. (2018) moderated this claim by observing poorer performances in patients when they were asked to integrate speech with complementary iconic gestures. Although these studies did not distinguish performances according to aphasia type, they do confirm the need for a precise qualification of the type of iconic gesture and its relationship to speech. Indeed, it appears that the advantage of a bimodal presentation is only present in the case of redundant and not complementary gestures. Eggenberger et al. (2016) have also proposed that future studies take individual differences into account, particularly when studying clinical populations.

Another pathology presenting a heterogeneous profile of language deficits, and particularly a limited verbal comprehension, is Specific Language Impairment (SLI) (Evans and Brown, 2016). Because this disorder is characterized by the presence of a language impairment in the absence of non-verbal cognitive impairments (Botting et al., 2010), it is of particular interest for the investigation of co-speech gesture integration. Using a Speech/Gesture Integration task [a paradigm created by Cocks et al. (2009)], Botting et al. (2010) not only highlighted poorer performances for SLI children, but also showed that they made more gesture-based errors. This would suggest that these children, although they did recognize hand movements, were unable to either extract the meaning from the gestures, or integrate it with the sentence context (Botting et al., 2010). These findings were later replicated by Wray et al. (2016), even after controlling for non-verbal cognition abilities. The difficulty for SLI children to integrate gesture meaning into a sentence context is consistent with language studies showing difficulties in integrating contextual information (Botting and Adams, 2005; Ryder et al., 2008). However, using a different paradigm, a study by Perrault et al. (2019) showed better performances when children with language disorders were faced with iconic gestures compared to typically developing (TD) children and children with autism spectrum disorders (ASD). Interestingly, the gestures in this study were devoid of sound and did not require any form or contextual integration to be understood; co-speech gestures used in Botting and Adams (2005) and Wray et al. (2016) studies were complementary to speech, while Perrault et al. (2019) used gestures *in place of* speech. More recently, Vogt and Kauschke (2017) highlighted a beneficial effect of bimodal iconic gesture presentation on word learning for SLI compared to TD children. These apparent contradictory results can be explained by the presentation format of the stimuli. Vogt and Kauschke (2017) presented face-to-face gestures while the previously discussed studies with null effects presented video clips (Botting et al., 2010; Wray et al., 2016). As has already been highlighted in this review, real-life gesture presentation has shown to be more efficient in improving comprehension (Holler et al., 2009; Sekine et al., 2015).

Perrault et al. (2019) also investigated gesture comprehension among ASD children. These children performed worse than TD and SLI children for co-speech gestures. The results were partly comparable to those of Dimitrova et al. (2017). These authors showed that compared to complementary co-speech gestures (for which performances were indeed poorer), redundant gestures improved performances for ASD children. They also demonstrated that gesture comprehension was linked to receptive language abilities.

Finally, several studies focused on the perception of gestures in patients with schizophrenia. An older study has shown a general impairment of gesture recognition (Berndl et al., 1986). However, there are numerous types of gestures, none of which are entirely processed in the same manner. In fact, patients present an inability to understand the meanings being metaphors or abstract concepts (Kircher et al., 2007). This is consistent with recent studies suggesting that recognition of metaphorical compared to iconic gestures is selectively impaired

(Straube et al., 2013, 2014; Nagels et al., 2019). Although iconic gesture recognition appears to be preserved, studies investigating the neural processes involved yield some interesting results (Straube et al., 2013, 2014; Schülke and Straube, 2019). Straube et al. (2013) found a disturbance in the activation of the left pMTG/STS and IFG for metaphorical gestures. A subsequent study specified the existence of a negative correlation between positive symptoms of schizophrenia and connectivity between the left IFG and left STS (Straube et al., 2014). In contrast, the activation of the left STS (and its connectivity to the left IFG and left MTG) for iconic gestures was comparable to that of healthy participants (Straube et al., 2014). Using tDCS, a recent study showed improved performances on a semantic-relatedness task when stimulating the frontal and fronto-parietal regions (Schülke and Straube, 2019). In schizophrenia, it would, therefore, appear that gesture recognition is selectively impaired for metaphorical gestures and preserved for iconic gestures. However, all these studies presented redundant iconic gestures. Given (1) the apparent distinction in processing redundant vs. complementary information and (2) the processing similarities for metaphorical and complementary iconic gestures (both involving IFG unification processes), contrasting these two types of gestures could be an avenue for further research.

In summary, although data from clinical studies presents a somewhat confusing picture, this can be resolved by considering the relationship between iconic gestures and the co-occurring speech as well as individual differences. Results for aphasic patients indeed suggest a differentiated effect of iconic gestures on comprehension, with redundant iconic gestures improving it and complementary iconic gestures hindering it. Studies on SLI reveal a positive effect of iconic gestures on comprehension, but only when these are presented face-to-face. A clear distinction of performance between complementary and redundant iconic gestures is found with ASD children as only the presence of the latter enhances comprehension. Finally, iconic gesture comprehension seems to be preserved in schizophrenia. Having said that, existing studies have predominantly focused on investigating redundant iconic gestures. Exploring the comprehension of complementary iconic gestures would consequently allow to paint a more complete picture of gesture-speech comprehension in schizophrenia.

CONCLUSION

Following this overview on the investigations of gesture-speech integration and the role of iconic gestures in language comprehension, an undeniable observation is the diversity of methods used, and the associated variation in results. Studies investigating neurologically intact individuals agree on attributing an active role to iconic gestures in improving language comprehension, particularly in an unfavorable listening context. But this does not imply that the gesture-speech integration is carried out in an automatic fashion and/or stems from a "unique integrated system" (McNeill, 1992; Kelly et al., 2010b). Rather, behavioral, electrophysiological and clinical studies results appear to plead in favor of the existence of two distinct systems, one being able to compensate the other in the event of an impairment (Perrault et al., 2019). Behavioral results suggest that gesture-speech integration can be modulated by the semantic overlap between both modalities (Holle and Gunter, 2007), intentionality (Kelly et al., 2007), or more generally in the presence of situational factors (Holle and Gunter, 2007). In other words, the automaticity of gesture-speech integration can be affected when attention is explicitly directed toward integration or when the task requires a controlled cognitive process, such as a lexical or semantic decision (Kelly et al., 2010b).

In clinical studies, depending on the pathology, the authors have identified either a parallel impairment of the verbal and gestural channels (thereby supporting the "unique integrated system"), or a preservation of the gestural channel allowing for better performances when the verbal channel is impaired. Discrepancies can also be found within the same pathology, depending on the relationship between gesture and speech and the task involved. In aphasia and ASD, the automaticity of integration is consistent with the poorer performances in presence of complementary iconic gestures but undermined in presence of redundant gestures. In SLI, the essential variable seems to be the presentation format of the stimuli. Further research on gesture-speech integration in schizophrenia would be needed in order to be able to distinguish between performances depending on the type of iconic gestures involved. All things considered, this variation in methodology could thus explain the current discussion regarding the automatic nature of gesture-speech integration.

These methodological variations could also serve as framework to analyze brain activity data. This review has emphasized the necessity of future research investigating the co-activated neural networks underlying gesture-speech integration for various types of iconic gestures, and differentiating the redundant from the complementary iconic gestures. Relatively new to the field, TMS (as well as its combination with neuroimaging or electrophysiological techniques) could also offer an interesting perspective on areas involved in gesture-speech integration. The combination of techniques would allow for a more precise qualification of the role and timing of the different regions involved. This is particularly relevant in relation to exploring the co-activated neural network involved in the processing of iconic gestures integration, as has been highlighted in this article. TMS studies could also help to shed some light on the presence or absence of early effects observed in electrophysiological studies. However, as mentioned previously, special attention must be paid to precisely characterizing the type of iconic gesture used as well as its relationship to speech (i.e., whether it is redundant, complementary, or congruent).

Finally, studying the cognitive processes underlying gesture-speech integration could also enhance our understanding of the latter. However, as has been emphasized, merely investigating cognitive processes does not imply taking individual variability into account. Therefore, considering individual differences is essential for a correct understanding of what is involved in gesture-speech integration.

Following this overview, **Table 1** provides a non-comprehensive but extended summary illustrating the possible variables that can be manipulated in the investigation of gesture-speech integration with an example of a study for each element. Appendix A presents a detailed summary of these variables for the studies explored in this paper.

In conclusion, although iconic gestures convey useful information for the listener, the specificities of iconic gestural and linguistic information (such as the automaticity of integration, the relationship between gesture and speech, or the brain regions involved) open wide fields of possible research. On a theoretical level, qualifying iconic gestures as manual movements with a semantic relationship to the co-occurring speech does not allow for a complete understanding of the results presented in the various studies. There is a clear need for going further and systematically specifying the type of iconic gesture (action, shape, size, and position) used, its manner of presentation (video clips or face-to-face) and its relation to speech (redundant, complementary, and incongruent).

The automatic nature of gesture-speech integration remains an issue. However, the authors' observations throughout this review support the theory of a modulated automaticity, depending on iconic gesture type, semantic overlap between gesture and co-occurring speech, particularly in the case of redundant vs. complementary, and individual differences in cognition. Although all electrophysiological studies have highlighted the existence of a semantic integration of information, they do not agree on the temporality of this integration (studies finding early and/or only late components). This disagreement mainly appears to stem from the use of different materials (simple vs. complex or redundant vs. complementary iconic gestures). Investigating early and late effects using similar material could help to resolve this issue.

TABLE 1 | Types of possible investigation and possible dependent variables in the study of iconic gestures in comprehension.

Methodological aspects	Possible variations
Investigation	– Behavioral (Beattie and Shovelton, 2002). – Electrophysiology (Wu and Coulson, 2007a). – Transcranial magnetic stimulation (Zhao et al., 2018). – Transcranial direct current stimulation (Cohen-Maximov et al., 2015). – Functional magnetic resonance imaging (Holle et al., 2008). – Magneto encephalogram (Drijvers et al., 2018). – Eye tracking (Beattie et al., 2010).
Gesture-speech integration	– Implicit (Sekine et al., 2015) or Explicit (Perrault et al., 2019).
Task during stimuli presentation	– Passive observation (Habets et al., 2011). – Dual task paradigm (Wu and Coulson, 2014). – Lexical decision (So et al., 2013). – Attentional task (Green et al., 2009). – Target relatedness task (Ping et al., 2014). – Stroop-like task (Kelly et al., 2010a).
Attention during stimuli presentation	– Speech. – Gesture (Bohn et al., 2020). – Stimuli as a whole (Vogt and Kauschke, 2017). – Unrelated aspect (Kelly et al., 2010a).
Type of iconic gesture	– Action (Stanfield et al., 2013). – Physical attributes (Dick et al., 2009). – Position (Beattie and Shovelton, 2002). – Typical vs. Atypical (Dargue and Sweller, 2018b).
Gesture-speech relationship	– Redundant (Holler et al., 2009) or Complementary (Kelly et al., 2004). – Congruent vs. Incongruent (Wu and Coulson, 2005). – Presence vs. Absence (Iani and Bucciarelli, 2017).
Type of stimuli	– Video clips (speech + gesture) (Kelly et al., 2010b). – Soundless video clips (Novack et al., 2016) – Live gestures (Kartalkanat and Göksun, 2020). – Target words/pictures (Bernardis et al., 2008; Wray et al., 2016). – Cartoons (Wu and Coulson, 2005).
Stimuli content	– Sentences (Momsen et al., 2020). – Single words (Sekine et al., 2020). – Narration (Macoun and Sweller, 2016). – Visual stimuli (Aussems and Kita, 2019).
Gesture length	– Full gesture (Kelly et al., 2007) or Stroke (So et al., 2013).
Origin of gesture	– Spontaneous (Holle et al., 2010) or Scripted (Holler et al., 2015).
Visibility of actor	– Knees up (Wolf et al., 2017). – Waist up (Dick et al., 2014). – Torso (Zhao et al., 2018). – Visible (Green et al., 2009) or Masked face (Zhao et al., 2018).

Brain imaging studies have facilitated a deeper analysis, while showing the involvement of the left IFG, pSTS, and MTG in gesture-speech integration. The variations in results are consistent with the authors' observation of the existence of different neural processing depending on the relationship between iconic gestures and speech. While existing studies have started to investigate the neural network involved in the processing of pantomimes, further research should explore the neural networks involved in the understanding of iconic gestures. As pointed out, brain imaging studies are particularly sensitive to the type of iconic gesture as well as the relationship it entertains with speech. Hence, future research could specify this information to make valid comparisons between studies as well as identifying the networks involved in different types of iconic gestures. Finally, since gesture-speech integration is a relatively recent field of investigation, studying the cognitive processes involved, such as working memory or attention, could allow for a better understanding of this integration.

AUTHOR CONTRIBUTIONS

KK: conceptualization and writing and editing – original draft preparation. IL and LL: supervision – reviewing and editing. WB: reviewing. MR: supervision – reviewing. All authors contributed to the article and approved the submitted version.

FUNDING

KK was funded by a doctoral award from the University of Mons.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.634074/full#supplementary-material

REFERENCES

Alfred, K. L., and Kraemer, D. J. (2017). Verbal and visual cognition: individual differences in the lab, in the brain, and in the classroom. *Dev. Neuropsychol.* 42, 507–520. doi: 10.1080/87565641.2017.1401075

Aussems, S., and Kita, S. (2019). Seeing iconic gestures while encoding events facilitates children's memory of these events. *Child Dev.* 90, 1123–1137. doi: 10.1111/cdev.12988

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., and Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47, 907–918. doi: 10.1016/j.neuron.2005.07.023

Başar, E. (2013). A review of gamma oscillations in healthy subjects and in cognitive impairment. *Int. J. Psychophysiol.* 90, 99–117. doi: 10.1016/j.ijpsycho.2013.07.005

Beattie, G., and Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: a semantic feature approach. *Gesture* 1, 129–149. doi: 10.1075/gest.1.2.03bea

Beattie, G., and Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *Br. J. Soc. Psychol.* 41, 403–417. doi: 10.1348/014466602760344287

Beattie, G., Webster, K., and Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *J. Lang. Soc. Psychol.* 29, 194–213. doi: 10.1177/0261927x09359589

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004a). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7, 1190–1192. doi: 10.1038/nn1333

Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004b). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/s0896-6273(04)00070-4

Bernardis, P., Salillas, E., and Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognit. Neuropsychol.* 25, 1114–1128. doi: 10.1080/02643290801921707

Berndl, K., Von Cranach, M., and Grüsser, O.-J. (1986). Impairment of perception and recognition of faces, mimic expression and gestures in schizophrenic patients. *Eur. Arch. Psych. Neurol. Sci.* 235, 282–291. doi: 10.1007/bf00515915

Bohn, M., Kordt, C., Braun, M., Call, J., and Tomasello, M. (2020). Learning novel skills from iconic gestures: a developmental and evolutionary perspective. *Psychol. Sci.* 31, 873–880. doi: 10.1177/0956797620921519

Botting, N., and Adams, C. (2005). Semantic and inferencing abilities in children with communication disorders. *Int. J. Lang. Commun. Disord.* 40, 49–66. doi: 10.1080/13682820410001723390

Botting, N., Riches, N., Gaynor, M., and Morgan, G. (2010). Gesture production and comprehension in children with specific language impairment. *Br. J. Dev. Psychol.* 28, 51–69. doi: 10.1348/026151009x482642

Cassell, J., McNeill, D., and McCullough, K.-E. (1999). Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics Cognit.* 7, 1–34. doi: 10.1075/pc.7.1.03cas

Church, R. B., and Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23, 43–71. doi: 10.1016/0010-0277(86)90053-3

Cocks, N., Byrne, S., Pritchard, M., Morgan, G., and Dipper, L. (2018). Integration of speech and gesture in aphasia. *Int. J. Lang. Commun. Disord.* 53, 584–591. doi: 10.1111/1460-6984.12372

Cocks, N., Morgan, G., and Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture* 11, 24–39. doi: 10.1075/gest.11.1.02coc

Cocks, N., Sautin, L., Kita, S., Morgan, G., and Zlotowitz, S. (2009). Gesture and speech integration: an exploratory study of a man with aphasia. *Int. J. Lang. Commun. Disord.* 44, 795–804. doi: 10.1080/13682820802256965

Cohen-Maximov, T., Avirame, K., Flöel, A., and Lavidor, M. (2015). Modulation of gestural-verbal semantic integration by tDCS. *Brain Stimul.* 8, 493–498. doi: 10.1016/j.brs.2014.12.001

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., and Wise, R. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain* 126, 1193–1201. doi: 10.1093/brain/awg104

Dahl, T. I., and Ludvigsen, S. (2014). How I see what you're saying: the role of gestures in native and foreign language listening comprehension. *Modern Lang. J.* 98, 813–833. doi: 10.1111/modl.12124

Dargue, N., and Sweller, N. (2018a). Donald Duck's garden: the effects of observing iconic reinforcing and contradictory gestures on narrative comprehension. *J. Exp. Child Psychol.* 175, 96–107. doi: 10.1016/j.jecp.2018.06.004

Dargue, N., and Sweller, N. (2018b). Not all gestures are created equal: the effects of typical and atypical iconic gestures on narrative comprehension. *J. Nonverbal Behav.* 42, 327–345. doi: 10.1007/s10919-018-0278-3

Dargue, N., and Sweller, N. (2020). Learning stories through gesture: gesture's effects on child and adult narrative comprehension. *Educ. Psychol. Rev.* 32, 249–276. doi: 10.1007/s10648-019-09505-0

Dargue, N., Sweller, N., and Jones, M. P. (2019). When our hands help us understand: a meta-analysis into the effects of gesture on comprehension. *Psychol. Bull.* 145:765. doi: 10.1037/bul0000202

De Lange, F. P., Jensen, O., Bauer, M., and Toni, I. (2008). Interactions between posterior gamma and frontal alpha/beta oscillations during imagined actions. *Front. Hum. Neurosci.* 2:7. doi: 10.3389/neuro.09.007

Demir-Lira, ÖE., Asaridou, S. S., Raja Beharelle, A., Holt, A. E., Goldin-Meadow, S., and Small, S. L. (2018). Functional neuroanatomy of gesture–speech integration in children varies with individual differences in gesture processing. *Dev. Sci.* 21:e12648. doi: 10.1111/desc.12648

Dick, A. S., Goldin-Meadow, S., Solodkin, A., and Small, S. L. (2012). Gesture in the developing brain. *Dev. Sci.* 15, 165–180. doi: 10.1111/j.1467-7687.2011.01100.x

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I, and Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum. Brain Map.* 30, 3509–3526. doi: 10.1002/hbm.20774

Dick, A. S., Mok, E. H., Beharelle, A. R., Goldin-Meadow, S., and Small, S. L. (2014). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Hum. Brain Map.* 35, 900–917. doi: 10.1002/hbm.22222

Dimitrova, N., Özçalı şkan, Ş, and Adamson, L. B. (2017). Do verbal children with autism comprehend gesture as readily as typically developing children? *J. Autism Dev. Disord.* 47, 3267–3280. doi: 10.1007/s10803-017-3243-9

Drijvers, L., and Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain Lang.* 177, 7–17. doi: 10.1016/j.bandl.2018.01.003

Drijvers, L., and Özyürek, A. (2020). Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. *Lang. Speech*. 63, 209–220. doi: 10.1177/0023830919831311

Drijvers, L., Özyürek, A., and Jensen, O. (2018). Hearing and seeing meaning in noise: alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. *Hum. Brain Map.* 39, 2075–2087. doi: 10.1002/hbm.23987

Drijvers, L., Vaitonytë, J., and Özyürek, A. (2019). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cognit. Sci.* 43:e12789.

Eggenberger, N., Preisig, B. C., Schumacher, R., Hopfner, S., Vanbellingen, T., Nyffeler, T., et al. (2016). Comprehension of co-speech gestures in aphasic patients: an eye movement study. *PloS one*. 11:e0146583. doi: 10.1371/journal.pone.0146583

Evans, J. L., and Brown, T. T. (2016). *Specific language impairment*, in *Neurobiology of language*. Netherlands: Elsevier, 899–912.

Eysenck, M. (2014). *Fundamentals of psychology*. London: Psychology Press.

Ferro, J. M., Santos, M. E., Castro-Caldas, A., and Mariano, M. G. (1980). Gesture recognition in aphasia. *J. Clin. Exp. Neuropsychol.* 2, 277–292. doi: 10.1080/01688638008403800

Fitzgibbon, S., Pope, K., Mackenzie, L., Clark, C., and Willoughby, J. (2004). Cognitive tasks augment gamma EEG power. *Clin. Neurophysiol.* 115, 1802–1809. doi: 10.1016/j.clinph.2004.03.009

Gainotti, G., and Lemmo, M. A. (1976). Comprehension of symbolic gestures in aphasia. *Brain Lang.* 3, 451–460. doi: 10.1016/0093-934x(76)90039-0

Glasser, M. L., Williamson, R. A., and Özçalı şkan, Ş (2018). Do children understand iconic gestures about events as early as iconic gestures about entities? *J. Psychol. Res.* 47, 741–754. doi: 10.1007/s10936-017-9550-7

Gough, P. M., Nobre, A. C., and Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci.* 25, 8010–8016. doi: 10.1523/jneurosci.2307-05.2005

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Map.* 30, 3309–3324. doi: 10.1002/hbm.20753

Habets, B., Kita, S., Shao, Z., Özyurek, A., and Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *J. Cognit. Neurosci.* 23, 1845–1854. doi: 10.1162/jocn.2010.21462

Hadar, U., and Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica* 115, 147–172.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Front. Psychol.* 4:416. doi: 10.3389/fpsyg.2013.00416

Hagoort, P., Baggio, G., and Willems, R. M. (2009). *Semantic unification*, in *The cognitive neurosciences*, 4th Edn. Cambridge: MIT press, 819–836.

Hald, L. A., Bastiaansen, M. C., and Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain Lang.* 96, 90–105. doi: 10.1016/j.bandl.2005.06.007

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature* 406, 147–150.

He, Y., Steines, M., Sommer, J., Gebhardt, H., Nagels, A., Sammer, G., et al. (2018). Spatial–temporal dynamics of gesture–speech integration: a simultaneous EEG-fMRI study. *Brain Struct. Funct.* 223, 3073–3089. doi: 10.1007/s00429-018-1674-5

Hein, G., and Knight, R. T. (2008). Superior temporal sulcus—it's my area: or is it? *J. Cognit. Neurosci.* 20, 2125–2136. doi: 10.1162/jocn.2008.20148

Hillyard, S. A., and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci.* 95, 781–787.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology* 30, 47–61. doi: 10.1111/j.1469-8986.1993.tb03204.x

Holle, H., and Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *J. Cognit. Neurosci.* 19, 1175–1192. doi: 10.1162/jocn.2007.19.7.1175

Holle, H., Gunter, T. C., Rüschemeyer, S.-A., Hennenlotter, A., and Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage* 39, 2010–2024. doi: 10.1016/j.neuroimage.2007.10.055

Holle, H., Obleser, J., Rueschemeyer, S.-A., and Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49, 875–884. doi: 10.1016/j.neuroimage.2009.08.058

Holler, J., and Beattie, G. (2003). How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process? *Semiotica* 146, 81–116.

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S. D., and Özyürek, A. (2015). Eye'm talking to you: speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Soc. Cognit. Affect. Neurosci.* 10, 255–261. doi: 10.1093/scan/nsu047

Holler, J., Shovelton, H., and Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *J. Nonverbal Behav.* 33, 73–88. doi: 10.1007/s10919-008-0063-9

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage* 36, 924–932. doi: 10.1016/j.neuroimage.2007.03.059

Iani, F., and Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *J. Mem. Lang.* 96, 110–121. doi: 10.1016/j.jml.2017.05.004

Jarrold, C., and Towse, J. N. (2006). Individual differences in working memory. *Neuroscience* 139, 39–50.

Jensen, O., Kaiser, J., and Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci.* 30, 317–324. doi: 10.1016/j.tins.2007.05.001

Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186

Kandana Arachchige, K. G., Holle, H., Loureiro, I. S., Blekic, W., Rossignol, M., and Lefebvre, L. (2018). The effect of verbal working memory load in speech/gesture integration processing. *Front. Neurosci.* 12:43. doi: 10.3389/conf.fnins.2018.95.00043

Kartalkanat, H., and Göksun, T. (2020). The effects of observing different gestures during storytelling on the recall of path and event information in 5-year-olds and adults. *J. Exp. Child Psychol.* 189:104725. doi: 10.1016/j.jecp.2019.104725

Kelly, S. D., Creigh, P., and Bartolotti, J. (2010a). Integrating speech and iconic gestures in a stroop-like task: evidence for automatic processing. *J. Cognit. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254

Kelly, S. D., Kravitz, C., and Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* 89, 253–260. doi: 10.1016/s0093-934x(03)00335-3

Kelly, S. D., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Lang. Cognit. Proc.* 24, 313–334. doi: 10.4324/9781003059783-8

Kelly, S. D., Özyürek, A., and Maris, E. (2010b). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327

Kelly, S. D., Ward, S., Creigh, P., and Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain Lang.* 101, 222–233. doi: 10.1016/j.bandl.2006.07.008

Kircher, T. T., Leube, D. T., Erb, M., Grodd, W., and Rapp, A. M. (2007). Neural correlates of metaphor processing in schizophrenia. *Neuroimage* 34, 281–289. doi: 10.1016/j.neuroimage.2006.08.044

Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal: evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/s0749-596x(02)00505-3

Krauss, R. M., Chen, Y., and Chawla, P. (1996). "Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us?," in *Advances in experimental social psychology*, ed. M. P. Zanna (Netherlands: Elsevier), 389–450. doi: 10.1016/s0065-2601(08)60241-5

Krauss, R. M., Morrel-Samuels, P., and Colasante, C. (1991). Do conversational hand gestures communicate? *J. Personal. Soc. Psychol.* 61:743. doi: 10.1037/0022-3514.61.5.743

Kuperberg, G. R., Sitnikova, T., and Lakshmanan, B. M. (2008). Neuroanatomical distinctions within the semantic system during sentence comprehension: evidence from functional magnetic resonance imaging. *Neuroimage* 40, 367–388. doi: 10.1016/j.neuroimage.2007.10.009

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*. 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., Van Petten, C. K., and Kluender, R. (2006). *Psycholinguistics electrified II (1994–2005)*, in *Handbook of psycholinguistics*. Netherlands: Elsevier, 659–724.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge: MIT press.

Luck, S. J., and Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components*. Oxford, UK: Oxford university press.

Luck, S. J., Woodman, G. F., and Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cognit. Sci.* 4, 432–440.

Macoun, A., and Sweller, N. (2016). Listening and watching: the effects of observing gesture on preschoolers' narrative comprehension. *Cognit. Dev.* 40, 68–81. doi: 10.1016/j.cogdev.2016.08.005

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology* 32, 4–18. doi: 10.1111/j.1469-8986.1995.tb03400.x

Margiotoudi, K., Kelly, S., and Vatakis, A. (2014). Audiovisual temporal integration of speech and gesture. *Procedia-Soc. Behav. Sci.* 126, 154–155. doi: 10.1016/j.sbspro.2014.02.351

Marslen-Wilson, W. D. (1984). "Function and process in spoken word recognition: A tutorial review," in *Attention and performance: Control of language processes*, eds H. Bouma and D. G. Bouwhuis (Hillsdale, NJ: Lawrence Erlbaum), 125–150.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, USA: University of Chicago press.

Miyake, H., and Sugimura, S. (2018). The effect of directive words on integrated comprehension of speech and iconic gestures for actions in young children. *Infant Child Dev*. 27:e2096. doi: 10.1002/icd.2096

Momsen, J., Gordon, J., Wu, Y. C., and Coulson, S. (2020). Verbal working memory and co-speech gesture processing. *Brain Cognit.* 146:105640. doi: 10.1016/j.bandc.2020.105640

Momsen, J., Gordon, J., Wu, Y. C., and Coulson, S. (2021). Event related spectral perturbations of gesture congruity: visuospatial resources are recruited for multimodal discourse comprehension. *Brain Lang.* 216:104916. doi: 10.1016/j.bandl.2021.104916

Nagels, A., Kircher, T., Grosvald, M., Steines, M., and Straube, B. (2019). Evidence for gesture-speech mismatch detection impairments in schizophrenia. *Psych. Res.* 273, 15–21. doi: 10.1016/j.psychres.2018.12.107

Nieuwland, M., Barr, D., Bartolozzi, B.-M., Darley, D., Ferguson, H., and Huettig, H. (2020). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *Phil. Transac. R. Soc. B Biol. Sci.* 375:20180522. doi: 10.1098/rstb.2018.0522

Novack, M. A., Wakefield, E. M., and Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition* 146, 339–348. doi: 10.1016/j.cognition.2015.10.014

Obermeier, C., and Gunter, T. C. (2014). Multisensory integration: the case of a time window of gesture–speech integration. *J. Cognit. Neurosci.* 27, 292–307. doi: 10.1162/jocn_a_00688

Özer, D., and Göksun, T. (2020a). Gesture use and processing: a review on individual differences in cognitive resources. *Front. Psychol.* 11:573555. doi: 10.3389/fpsyg.2020.573555

Özer, D., and Göksun, T. (2020b). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Lang. Cognit. Neurosci.* 35, 896–914. doi: 10.1080/23273798.2019.1703016

Özyürek, A., Willems, R. M., Kita, S., and Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cognit. Neurosci.* 19, 605–616. doi: 10.1162/jocn.2007.19.4.605

Pascual-Leone, A., Walsh, V., and Rothwell, J. (2000). Transcranial magnetic stimulation in cognitive neuroscience–virtual lesion, chronometry, and functional connectivity. *Curr. Opin. Neurobiol.* 10, 232–237. doi: 10.1016/s0959-4388(00)00081-7

Perrault, A., Chaby, L., Bigouret, F., Oppetit, A., Cohen, D., Plaza, M., et al. (2019). Comprehension of conventional gestures in typical children, children with autism spectrum disorders and children with language disorders. *Neuropsychiatrie de l'Enfance et de l'Adolescence*. 67, 1–9. doi: 10.1016/j.neurenf.2018.03.002

Ping, R. M., Goldin-Meadow, S., and Beilock, S. L. (2014). Understanding gesture: is the listener's motor system involved? *J. Exp. Psychol. General* 143:195. doi: 10.1037/a0032246

Preisig, B. C., Eggenberger, N., Cazzoli, D., Nyffeler, T., Gutbrod, K., Annoni, J.-M., et al. (2018). Multimodal communication in aphasia: perception and production of co-speech gestures during face-to-face conversation. *Front. Hum. Neurosci.* 12:200. doi: 10.3389/fnhum.2018.00200

Price, C., Indefrey, P., and van Turennout, M. (1999). "The neural architecture underlying the processing of written and spoken word forms," in *The neurocognition of language*, eds C. M. Brown and P. Hagoort (Oxford: Oxford Scholarship Online), 211–240.

Quandt, L. C., Marshall, P. J., Shipley, T. F., Beilock, S. L., and Goldin-Meadow, S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: an EEG study of action production and gesture observation. *Neuropsychol.* 50, 2745–2751. doi: 10.1016/j.neuropsychologia.2012.08.005

Rossini, P. M., and Rossi, S. (2007). Transcranial magnetic stimulation: diagnostic, therapeutic, and research potential. *Neurology* 68, 484–488. doi: 10.1212/01.wnl.0000250268.13789.b2

Ryder, N., Leinonen, E., and Schulz, J. (2008). Cognitive approach to assessing pragmatic language comprehension in children with specific language impairment. *Int. J. Lang. Commun. Disord.* 43, 427–447. doi: 10.1080/13682820701633207

Schubotz, L., Özyürek, A., and Holler, J. (2019). Age-related differences in multimodal recipient design: younger, but not older adults, adapt speech and co-speech gestures to common ground. *Lang. Cognit. Neurosci.* 34, 254–271. doi: 10.1080/23273798.2018.1527377

Schülke, R., and Straube, B. (2019). Transcranial direct current stimulation improves semantic speech–gesture matching in patients with schizophrenia spectrum disorder. *Schizophrenia Bull.* 45, 522–530. doi: 10.1093/schbul/sby144

Sekine, K., Schoechl, C., Mulder, K., Holler, J., Kelly, S., Furman, R., et al. (2020). Evidence for children's online integration of simultaneous information from speech and iconic gestures: an ERP study. *Lang. Cognit. Neurosci.* 35, 1283–1294. doi: 10.1080/23273798.2020.1737719

Sekine, K., Sowden, H., and Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cognit. Sci.* 39, 1855–1880. doi: 10.1111/cogs.12221

Sitnikova, T., Kuperberg, G., and Holcomb, P. J. (2003). Semantic integration in videos of real–world events: an electrophysiological investigation. *Psychophysiology* 40, 160–164. doi: 10.1111/1469-8986.00016

So, W. C., Low, A., Yap, D. F., Kheng, E., and Yap, M. (2013). Iconic gestures prime words: comparison of priming effects when gestures are presented alone and when they are accompanying speech. *Front. Psychol.* 4:779. doi: 10.3389/fpsyg.2013.00779

Srinivasan, R. (2005). "High-resolution EEG: theory and practice," in *Event-related potentials: A methods handbook*, ed. T. C. Handy (Cambridge: MIT Press), 167–188.

Stanfield, C., Williamson, R., and Özçalışkan, S. (2013). How early do children understand gesture-speech combinations with iconic gestures? *J. Child Lang.* 41, 462–471. doi: 10.1017/s0305000913000019

Straube, B., Green, A., Bromberger, B., and Kircher, T. (2011). The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum. Brain Map.* 32, 520–533. doi: 10.1002/hbm.21041

Straube, B., Green, A., Sass, K., and Kircher, T. (2014). Superior temporal sulcus disconnectivity during processing of metaphoric gestures in schizophrenia. *Schizophrenia Bull.* 40, 936–944. doi: 10.1093/schbul/sbt110

Straube, B., Green, A., Sass, K., Kirner-Veselinovic, A., and Kircher, T. (2013). Neural integration of speech and gesture in schizophrenia: evidence for differential processing of metaphoric gestures. *Hum. Brain Map.* 34, 1696–1712. doi: 10.1002/hbm.22015

Straube, B., Wroblewski, A., Jansen, A., and He, Y. (2018). The connectivity signature of co-speech gesture integration: the superior temporal sulcus modulates connectivity between areas related to visual gesture and auditory speech processing. *NeuroImage* 181, 539–549. doi: 10.1016/j.neuroimage.2018.07.037

Thompson, L. A., and Guzman, F. A. (1999). Some limits on encoding visible speech and gestures using a dichotic shadowing task. *J. Gerontol.Ser. B Psychol. Sci. Soc. Sci.* 54, 347–349.

Vogt, S., and Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *J. Child Lang.* 44, 1458–1484. doi: 10.1017/s0305000916000647

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., and Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338.

Wang, X. J. (2003). "Neural oscillations," in *Encyclopedia of cognitive science*, ed. L. Nadel (London: MacMillan), 272–280.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cognit. Sci.* 7, 553–559. doi: 10.1016/j.tics.2003.10.012

Weiss, S., and Mueller, H. M. (2012). "Too many betas do not spoil the broth": the role of beta brain oscillations in language processing. *Front. Psychol.* 3:201. doi: 10.3389/fpsyg.2012.00201

Willems, R. M., Özyürek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cerebral Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Willems, R. M., Özyürek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47, 1992–2004. doi: 10.1016/j.neuroimage.2009.05.066

Wolf, D., Rekittke, L.-M., Mittelberg, I., Klasen, M., and Mathiak, K. (2017). Perceived conventionality in co-speech gestures involves the fronto-temporal language network. *Front. Hum. Neurosci.* 11:573. doi: 10.3389/fnhum.2017.00573

Wray, C., Norbury, C. F., and Alcock, K. (2016). Gestural abilities of children with specific language impairment. *Int. J. Lang. Commun. Disord.* 51, 174–182. doi: 10.1111/1460-6984.12196

Wu, Y. C., and Coulson, S. (2005). Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology*. 42, 654–667. doi: 10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., and Coulson, S. (2007a). How iconic gestures enhance communication: an ERP study. *Brain Lang.* 101, 234–245. doi: 10.1016/j.bandl.2006.12.003

Wu, Y. C., and Coulson, S. (2007b). Iconic gestures prime related concepts: an ERP study. *Psychon. Bull. Rev.* 14, 57–63. doi: 10.3758/bf03194028

Wu, Y. C., and Coulson, S. (2010). Gestures modulate speech processing early in utterances. *Neuroreport* 21:522. doi: 10.1097/wnr.0b013e32833904bb

Wu, Y. C., and Coulson, S. (2014). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychol.* 153, 39–50. doi: 10.1016/j.actpsy.2014.09.002

Zhao, W., Riggs, K., Schindler, I., and Holle, H. (2018). Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration. *J. Neurosci.* 38, 1891–1900. doi: 10.1523/jneurosci.1748-17.2017

Zhu, Z., Hagoort, P., Zhang, J. X., Feng, G., Chen, H.-C., Bastiaansen, M., et al. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *Neuroimage* 60, 2230–2237. doi: 10.1016/j.neuroimage.2012.02.036

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership