Speech disfluencies and hand gestures as metacognitive cues

Begüm Yılmaz[1], Reyhan Furman[2], Tilbe Göksun[1*], Terry Eskenazi[1*]

[1]Department of Psychology, Koç University
[2]School of Psychology and Humanities, University of Central Lancashire
*These authors made equal contributions to this manuscript.

Author Note

Begüm Yılmaz https://orcid.org/0009-0005-7027-8846

Reyhan Furman https://orcid.org/0000-0001-6034-3820

Tilbe Göksun https://orcid.org/0000-0002-0190-7988

Terry Eskenazi https://orcid.org/0000-0001-6777-0753

Correspondence regarding the article should be send to Begüm Yılmaz, Rumelifeneri, Sarıyer Rumeli Feneri Yolu, 34450 Sarıyer/İstanbul, Email: byilmaz21@ku.edu.tr

CRediT Authorship contribution statement

Begüm Yılmaz: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - editing & reviewing. **Reyhan Furman:** Conceptualization, Writing - editing & reviewing. **Tilbe Göksun:** Conceptualization, Methodology, Writing - editing & reviewing, Funding acquisition, Project administration, Supervision. **Terry Eskenazi:** Conceptualization, Methodology, Writing - editing & reviewing, Project administration, Supervision.

Acknowledgement

This study has been supported by the James S. McDonnell Foundation Scholar Award (220020510) given to Tilbe Göksun. We thank Emel Nur Kaya, Sultan Karakaş, Ecesu Özcan, and İlhan Emre Yavuz for their help in the data coding process.

Declaration of interest

None of the authors of this study has any conflict of interest.

Abstract

How language interacts with metacognitive processes is an understudied area. Earlier research shows that people produce disfluencies (i.e., "*uh*" s or "*um*" s) in their speech when they are not sure of their answers, indicating metacognitive monitoring. Gestures have monitoring and predictive roles in language, also implicating metacognitive processes. Further, the rate of speech disfluencies and gestures change as a function of the communicational setting. People produce fewer disfluencies and more gestures when they can see the listener than when the listener is not visible. In the current study, 50 participants (32 women, *Mage* = 21.16, *SD* = 1.46) were asked 40 general knowledge questions, either with a visible (*n* = 25) or nonvisible (*n* = 25) listener. They provided feelings-of-knowing (FOK) judgment immediately after seeing the question and were asked to think aloud while pondering their answers. Then, they provided retrospective confidence judgments (RCJs). Results showed that gestures and speech disfluencies were not related either to the accuracy or the FOK judgments. However, both gestures and speech disfluencies predicted RCJs uniquely and interactively. Speech disfluencies negatively predicted RCJs. In contrast, hand gestures were positively related to RCJs. Importantly, the use of gestures was more strongly related to RCJs when disfluencies were also higher. No effect of communicational setting on the rate of gestures or speech disfluencies was found. These results highlight the importance of multimodal language cues in elaboration of metacognitive judgments.

Keywords: Metacognition, retrospective confidence judgments, feelings-of-knowing, speech disfluency, gesture production.

Metacognitive processes involve the ability to reflect on and regulate one's own cognitive processes, serving a host of adaptive functions (Frith, 2012). Most research on metacognition is built upon visual perception and memory (Rouault et al., 2018; Palmer et al., 2014). Little is known about the ways in which metacognitive processes are linked with other cognitive processes, such as language. Here we focus on two aspects of language—speech disfluencies and co-speech gestures —with respect to their reciprocal interaction with metacognitive processes. Prior research shows that people who are less confident in their answers produce more speech disfluencies (Smith & Clark, 1993), suggesting a link between speech related cues and metacognitive evaluations. Co-speech gestures, on the other hand, may have a predictive function. For example, questions asked with gestures get faster responses (ter Bekke et al., 2024a), and gestures precede their lexical affiliates (ter Bekke et al., 2024b). Research also indicates that gestures might accompany disfluent speech (Arslan et al., 2022; 2024).

Furthermore, speech disfluencies and gestures also vary as a function of the communicative pressure. People speak less fluently when they cannot see the listener (Alibali et al., 2001; Kasl & Mahl, 1965; Oviatt, 1995; Rimé, 1982). They also use fewer gestures when the listener is not visible (Bavelas et al., 2002; Cohen & Harrison, 1973; Krauss et al., 1995). These findings indicate that communicative pressure can affect how speakers modulate their speech and gestures. Earlier studies investigating metacognitive processes in speech production use nonvisible listeners to prevent any visual cues from affecting participants' responses (Smith & Clark, 1993; Swerts & Krahmer et al., 2005). In the current study, we specifically investigate the ways in which speech disfluencies and gestures interact with metacognitive processes, and how the rate of disfluent speech and gesture production change as a function of the communicational setting.

1.1 Metacognitive processes

Metacognition refers to thoughts about cognitive phenomena (Flavell, 1979). It plays a crucial role in monitoring and guiding cognitive processes, such as perception (Sherman et al., 2015), memory (Risko & Dunn, 2015), and decision-making (Double & Birney, 2017). Successful metacognition requires effective monitoring of one's first-order cognitive processes and controlling them to strategize behavior (Nelson, 1990). Researchers mostly investigate metacognitive processes through feelings-of-knowing (FOK) (Morson et al., 2015; Sacher et al., 2015) or retrospective confidence judgments (RCJs) (Mazor et al., 2022; Palmer et al., 2014). FOK judgements are prospective, asking for one's prediction of the future ability to recognize an answer. As such, FOK is elicited before one answers a question. In contrast, confidence judgments are given retrospectively after providing a response. Earlier studies show that these metacognitive judgments may tap into distinct processes (Fleming et al., 2016; Mazancieux et al., 2020). As such, FOK may arise as a result of cue familiarity or accessibility of the target, whereas RCJ reflects one's sensitivity to first-order performance (Chua & Solinger, 2015). Thus, it is crucial to take into account these different parameters while investigating how metacognition is linked with other cognitive processes such as language.

1.2 Speech disfluencies as metacognitive cues

Disfluency is defined as the subjective experience of difficulty in cognitive tasks, involving decision-making, visual perception, memory, and language (Diemand-Yauman et al. 2011; Fraundorf & Watson, 2014; Oppenheimer, 2008; Song & Schwarz, 2008; Yue et al., 2013). In a flawed process, speech production includes errors, pauses, or repetitions (Corley & Hartsuiker, 2003; Maclay & Osgood, 1959; Tree, 1995). For instance, people produce six disfluencies per 100 words in spoken English on average (Tree, 1995). Earlier findings show that speech disfluencies are not distributed arbitrarily in sentences, they occur when there is a low-frequency or unpredictable word (Beattie & Butterworth, 1979; Corley & Hartsuiker, 2003; Levelt, 1983; Schnadt & Corley, 2006), when the topic is unfamiliar (Merlo & Mansur, 2004), when there are multiple meaning words (Schachter et al., 1991; Schachter et al., 1994), or when there is uncertainty (Brennan & Williams, 1995; Smith & Clark, 1993; Swerts & Krahmer, 2005). A recent study also shows that listeners can make adaptations while processing disfluent speech, indicating that it can be a flexible cue based on the context it is used (Yoon & Brown-Schmidt, 2024).

Speech disfluencies may have multiple functions, serving cognitive or communicative purposes (Clark & Tree, 2002; Corley & Stewart, 2008; Levelt, 1983). From a cognitive perspective, individuals can produce disfluencies to edit their speech (Levelt, 1983). Nooteboom (1980) proposes the Main Interruption Rule, which suggests that speech is interrupted immediately upon detecting any trouble. Building on this, Levelt (1983) claims that the interruptions in speech production may be an upshot of metacognitive monitoring processes. As individuals speak, they can monitor their speech and be their own listeners, evaluating the speech and what they intend to say. If discrepancies arise, speakers can stop the speech flow and adjust their speech to align with the intention. This is what Levelt (1983) refers to as the perceptual theory of monitoring, which is claimed to play a significant role in speech production processes.

Speech disfluencies may also serve communicative functions. Clark and Tree (2002) claim that speech disfluencies, such as "uh"s or "um"s in speech, might serve as a word, to inform the listener that there is a delay in ongoing speech. Thus, disfluencies may act as a "time-buying" tool to remain on the conversational floor. Another view suggests that disfluencies are produced as a signal about the message conveyed (Corley & Stewart, 2008; Smith & Clark, 1993). Earlier research shows that when people are less certain of their answers, they produce more disfluencies (Smith & Clark, 1993; Swerts & Krahmer, 2005). People also rate those who produce disfluent speech as less knowledgeable (Brennan & Williams, 1995; Clark & Krych, 2004; Swerts & Krahmer, 2005).

These findings clearly suggest a bidirectional relationship between metacognitive processes and disfluencies. Disfluent speech might act as a metacognitive cue which triggers the system and signals that one is not sure of their answer. It is also possible that disfluencies are the by-products of metacognitive processes. That is, low confidence in one's process can manifest in disfluent speech, perhaps to allow one to slow down and correct speech errors. However, earlier studies examined the relation between metacognitive processes and speech disfluencies only through feelings-of-knowing judgements (FOKs) rather than retrospective confidence judgments (RCJs) (Brennan & Williams, 1995; Smith & Clark, 1993; Swerts & Krahmer, 2005). Since FOK judgements are given prospectively and they can be produced as a result of cue familiarity or target accessibility, it is not known whether RCJs are influenced by disfluencies in speech.

1.3 Co-speech gestures as metacognitive cues

Language is multimodal, including cues such as hand gestures, gaze, nodding, and shoulder movements (Bortfeld et al., 2001; Fröhlich et al., 2019; Swerts & Krahmer, 2005; Vigliocco et al., 2014). Among them, gestures as embodied cues have been shown to convey and encode meaning, having a complementary role in speech (McNeill, 1992; Kendon, 1988). McNeill (1992) proposes a coding scheme of co-speech gestures based on people's narrative discourse. Accordingly, gestures are categorized into four types, namely iconic, deictic (pointing), metaphoric, and beat gestures. Iconic gestures are closely related to the semantic content and refer to a concrete object, concept, or event (i.e., raising the hand(s) to refer to a high mountain). Metaphoric gestures, although resembling iconic gestures, refer to an abstract concept or an event (i.e., raising the hand(s) to refer to one’s high expectations). Along with these representational gestures, deictic gestures refer to pointings with one's index finger(s) or the entire hand(s), and are mainly used to refer to an item in the immediate environment. Besides, McNeill (1992) introduces beat gestures, which are the rhythmic hand movements accompanying speech that do not necessarily carry a specific meaning, thus are classified as nonrepresentational gestures (but see Ferrari & Hagoort, 2025). There also exist other nonreferential gestures, such as palm-up gestures (i.e., open hands towards up). These gestures have a pragmatic nature, indicating absence of knowledge, uncertainty, or interrogatives, and do not have a semantic relation with speech content (Cooperrider et al., 2018; Kendon, 2004).

Previous literature suggests a link between embodied motor actions and metacognitive processes (Alban & Kelley, 2013; Fleming et al., 2015; Koriat & Nussinson, 2009; Palser et al., 2018; Patel et al., 2012; Wokke et al., 2020). For example, when asked to contract eyebrows, people provided lower judgments-of-learning (JOL) on the items they studied compared to those who were asked to raise their eyebrows (Koriat & Nussinson, 2009). In another study, those who were less confident in their answers on an execution task moved their hands more slowly compared to those who had higher confidence, and this finding has also been extended to observing others actions as well (Patel et al., 2012). One proposed mechanism between embodied actions and the metacognitive system is the mental effort devoted to a task (Koriat et al., 2006; Koriat & Nussinson, 2009). Mental effort is distinguished as data-driven or goal-driven, and it is argued that these two drives can operate simultaneously (Koriat et al., 2006). In the former, people might put more effort on the task at hand based on its intrinsic difficulty as a bottom-up influence, resulting in lower metacognitive judgments. In contrast, goal-driven investment of effort is based on the idea that people can adjust how much effort they put in a task as a top-down fashion grounded on their willingness, resulting in higher estimations of performance. For instance, in Koriat and Nussinson (2009), asking participants to contract their eyebrows may have created the sense of intrinsic difficulty of the studied items, working as a bottom-up influence, consequently leading participants to provide lower judgments-of-learning (JOL). In contrast, when the same participants were asked to contract their eyebrows for the items they wanted to concentrate on under time pressure, their JOLs increased. This shows that top-down influence of intentionally allocating study time to favorable items while using the embodied cues have resulted in higher metacognitive judgments.

Gestures are embodied sensory-motor actions that help speakers express their thoughts, either with communicative or cognitive purposes (Kita et al., 2017; Goldin-Meadow, 1999). However, only a few studies examined whether hand gestures as embodied cues are related to metacognitive processes (Çapan et al., 2024; Ferrari & Hagoort, 2025). Gestures, like other embodied cues, can also influence metacognitive assessments of first-order performance within data-driven or goal-driven streams. People might produce gestures in a question answering paradigm as a result of online difficulty of the task at hand. In this case, gestures might be used as an expression of metacognitive states in the sense that people are not sure of their answers. Conversely, it is possible that gestures are produced to support mental resources to the problem at hand and put more effort intentionally. This might result in a metacognitive illusion that the person is sure of their answer. A recent study investigating metacognitive processes and gestures showed that encouraging gesture use boosts performance assessments in a spatial reasoning task (Çapan et al., 2024). In the study, participants who were prompted to use gestures while solving a mental rotation task reported higher confidence levels compared to those who did not receive the same prompt. However, this study investigated the relation between metacognitive processes and gestures through spatial reasoning or learning tasks manipulating gesture use, and the mechanism in which metacognitive processes interact with spontaneous co-speech gesture use in speech production remains unknown.

Gesture and speech are part of an integrated language system (Kita et al., 2017; McNeill, 1992). Supporting this, earlier studies show that gestures may accompany disfluent speech more than fluent speech (Finlayson et al., 2003; Arslan et al., 2022; 2024; but see Akhavan et al., 2016; Avcı et al., 2022; Cravotta et al., 2021; Hoetjes et al., 2014). Research also shows that when gesture use is restricted, people produce more speech disfluencies (Finlayson et al., 2003; Morsella & Krauss, 2004; Rauscher et al., 1996). It is less known whether speech disfluencies and gestures are uniquely and interactively related to metacognitive judgments. Koriat et al. (2006) argues that data-driven and goal-driven processes are not mutually exclusive but work in a dynamic fashion, combining top-down and bottom-up flows of information. Speech disfluencies might heighten cognitive demand, requiring the cognitive system to recruit additional resources such as hand gestures. In that case, the use of gestures when there is speech disfluency might be associated with online difficulty, resulting in lower confidence. On the other hand, if gestures are intentionally produced to help overcome speech errors, it might ease the speaking process itself, creating an additional sense of ease in the speaker. This feeling of ease as well as the willingness to correct speech errors might in turn have a positive effect on the metacognitive judgments, increasing confidence. So far, no study investigated the relation among hand gestures, speech disfluencies, and metacognitive judgments together, which is the main goal of the present work.

1.4 The effect of visibility of a listener on gesture and speech disfluency production

Previous studies investigating metacognitive processes and language involved a nonvisible listener (Brennan & Williams, 1995; Smith & Clark, 1993; Swerts & Krahmer, 2005); however, the presence or the visibility of a listener can affect the rate of both speech disfluency (Alibali et al., 2001; Kasl & Mahl, 1965; Lickiss & Wellens, 1978; Oviatt, 1995; Rimé et al., 1982) and gesture use (Alibali et al., 2001; Bavelas et al., 2002; 2008; Krauss et al., 1995).

Studies show that people speak less fluently when they cannot see the listener (Alibali et al., 2001; Kasl & Mahl, 1965; Oviatt, 1995; Rimé, 1982; but see Lickiss & Wellens 1978). That is, individuals produce more "uh"s or "um"s when the visibility of the listener is blocked, and only their voices can be heard. It is possible that when the listener is not visible, the speaker may not get help from non-verbal cues (i.e.,. eye gaze or nodding) as alternative channels to convey information; resulting in higher rates of disfluencies (Bortfeld et al., 2001; Brennan, 2002). Nevertheless, this conclusion is based on very few studies, necessitating further research.

There is also growing evidence that people use more gestures in dialogues compared to monologues (Bavelas et al., 2014a; see Bavelas & Healing, 2013 for a review). Besides, studies indicate that speakers gesture more when the listener is visible (Alibali et al., 2001; Cohen & Harrison, 1973; Krauss et al., 1995). That is, people might switch the communicational channel from an auditory (i.e., spoken language) to a visual one (i.e., hand gestures) when the listener is visible as they provide more information through gestures independent of the message verbally expressed. However, the effect of listener visibility on gesture use may not always be straightforward, which might depend on the task at hand or differentially affect the gesture types. For example, Alibali et al. (2001) showed that people produce more representational (i.e., iconic and metaphoric) gestures while retelling a story when the listener is visible; however, they could not find such an effect for nonrepresentational gestures (i.e., beats). In addition, Bavelas et al.

(2008) found that although the overall rate of gestures did not change depending on the visibility of the listener in a visual description task, participants used more informative gestures when their gestures were seen.

Other studies, however, do not show the effect of listener visibility on gesture production, highlighting the inconsistencies in the literature (Bavelas et al., 2014b; Pine et al., 2010). One possible reason for these inconsistencies is the way listener visibility is manipulated across studies. For example, in Pine et al. (2010), although a screen was used to block the listener's visibility, participants remained engaged in a dialogue in both the visible and non-visible listener conditions. Previous research shows that the effect of dialogue is higher than the effect of visibility on overall gesture production, potentially overriding the effect of visibility in studies with null findings (Bavelas et al., 2008). In addition, most studies investigate the relation between gestures and speech disfluencies using spatial tasks, thus it is critical to test their relationship with different tasks and in diverse contexts.

1.5 The present study

Our study investigated the metacognitive processes in disfluent speech production and gesture use. Previous research suggests that disfluencies can inform metacognitive monitoring reflecting in one's confidence (Smith & Clark, 1993; Swerts & Krahmer, 2005). Gestures might also play a predictive role in spoken language (Ferré, 2010; Seyfeddinipur, 2006; ter Bekke et al., 2024a, 2024b), suggesting its complementary role in speech processes. Earlier studies on the relation between metacognitive judgments and speech disfluencies use feelings-of-knowing (FOK) judgments, and it is not known how speech related cues are associated with retrospective confidence judgments (RCJs). Additionally, no study so far has investigated the interaction of metacognitive processes with speech (i.e., disfluent speech) and co-speech gestures in different communicational settings (i.e., visible vs nonvisible listener).

Speakers also produce gestures at a varying degree, suggesting individual differences in cognitive processes (Hostetter & Alibali, 2007; Özer & Göksun, 2020; Richmond et al., 2003). These individual differences can be related to one's working memory capacity (Marstaller & Burianová, 2013), spatial reasoning skills (Galati et al., 2018; Göksun et al., 2013), and even empathy (Canarslan & Chu, 2024; Chu et al., 2014). It is possible that those who have heightened awareness regarding hand gestures might use more gestures. Along with individual differences, task related factors such as question difficulty might also influence gesture production (Chu & Kita, 2011; Melinger & Kita, 2007; but see Sassenberg & Van Der Meer, 2010). Previous work shows that when the difficulty of problems increases (i.e., mental rotation task), people use more gestures (Chu & Kita, 2011). As the difficulty of questions increases, we know that the accuracy and metacognitive judgments might also decrease. Thus, it is crucial to take into account these sources of variability and how they may be interrelated in gesture production in the current study.

We addressed four main research questions: 1) Do speech disfluencies function as metacognitive cues informing confidence judgements? 2) Does the production of gestures relate to confidence judgments through metacognitive monitoring processes? 3) Do speech disfluencies and gestures interactively relate to metacognitive judgments? 4) How does the rate of disfluent speech and gesture production change as a function of the listener's visibility? Besides these four main questions, we also ask whether individual differences (i.e., gesture awareness) were related to gesture production. To answer these research questions, we created a paradigm in which individuals answer general knowledge questions and verbalize their thought process while answering the questions. Before verbalizing their thoughts, participants were asked to provide feelings-of-knowing (FOK) judgments immediately after they see the question. Then, they were shown the question with four response options and asked to verbalize their thought process. After they gave an answer, they were asked to provide a retrospective confidence judgment, rating how sure they are of their answer.

First, we predict that individuals who produced disfluent speech would provide lower confidence ratings as disfluencies might have triggered the metacognitive system when one does not know the answer. We expect this effect to be more pronounced in confidence judgments given retrospectively rather than in feelings-of-knowing (FOK) judgments. That is because disfluencies will have cued the speaker to monitor and evaluate their thought processes online as they unfold, which would be reflected retrospectively.

For the second research question, on gesture production and metacognitive judgments, we expect trial difficulty to be related to the use of gestures. In particular, we predict that more difficult questions would yield a higher number of gestures. Beyond that, if gestures are linked to inherent difficulty of the task, we expect those who produce more gestures to provide lower retrospective metacognitive judgments. This relation should be stronger for retrospective confidence ratings rather than FOK judgements given prospectively if gestures inform metacognitive evaluations, rather than the opposite. Alternatively, if gestures are produced as a result of monitoring as a way to allocate more cognitive resources to the task intentionally, then the use of gestures should boost confidence.

Regarding the third question, if producing more gestures while answering the questions reduces attention allocated to speech production and increases speech errors consequently, gestures might accompany disfluent speech more. In that situation, these embodied cues elicited as a bottom-up influence, due to intrinsic difficulty of the task, then it might be possible that the use of gestures when there is speech disfluency might amplify the effect of speech disfluencies on metacognitive judgments. Then, this might result in lower confidence ratings. However, it is also possible that the use of gestures when an error in speech detected might act as a compensatory mechanism, creating an illusory effect as if the speaker knows the answer, dampening the effect of speech disfluency on metacognitive judgments. Thus, we expect that if participants produce gestures along with disfluent speech to decrease speech errors, this might be reflected in metacognitive judgments. That is, when speech disfluencies and gestures are produced in a given trial, these gestures would compensate for speech difficulties, providing the participant the illusion that they know the answer, increasing their retrospective confidence.

For the fourth question, we hypothesize that contextual differences (i.e., the visibility of the listener) would affect speech disfluency and gesture rates. With a visible listener, we expect individuals to produce a higher number of gestures and lower rate of disfluencies. When a listener is visible, other paralinguistic cues (i.e., co-speech gestures) might act as alternative channels while sharing knowledge, easing up the speech production and thus reducing disfluency rates. Lastly, we expect that individual differences (i.e., gesture awareness) might affect gesture production rates. That is, participants with higher gesture awareness might produce more gestures.

Method

2.1 Participants

Data was collected from 54 participants; however, 4 participants were excluded from the analyses as three of them failed to follow instructions and equipment error occurred for one participant. The final sample consisted of 50 Turkish-speaking participants (32 women, *Mage* =

21.16, *SD* = 1.46). The sample size was determined based on previous research that examined related processes using similar materials and designs (Kimble & Seidel, 1991; Smith & Clark, 1993). Participants were recruited from [blinded] University subject pool and given either course credit or compensated with 50₺ for their participation—those who were older than 18 years old and native Turkish speakers were included in the study. This study has been reviewed and approved by [blinded] University Committee on Human Research under the code of 2023.323.IRB3.143.

2.2 Measurements

2.2.1 General knowledge task

There were 40 general knowledge questions with 4-alternative-forced-choice (4AFC) and the content of speech was controlled by asking all participants the same set of questions. The reason for the 4AFC design was to encourage participants not to merely provide one-word answers but to vocalize their thought process, allowing them to produce speech disfluencies and co-speech gestures at a measurable rate. The questions and answers were randomized to ensure that each participant received the questions and the four possible responses in a random order. To determine the question difficulty, we pretested another group of participants (N = 20) with 50 general knowledge questions. We then eliminated the questions correctly answered by more than 85% of the participants, which made up 10 questions in total (*Mcorrect* = 18, *SD* = 0.94), leaving us with 40 questions. This helped prevent participants from producing one-word or short answers. For the remaining 40 questions (*Mcorrect* = 11.03, *SD* = 3.96), we determined question difficulty (1:very easy, 4:very difficult) for each question based on how many participants answered them correctly.

2.2.2 Gesture awareness scale

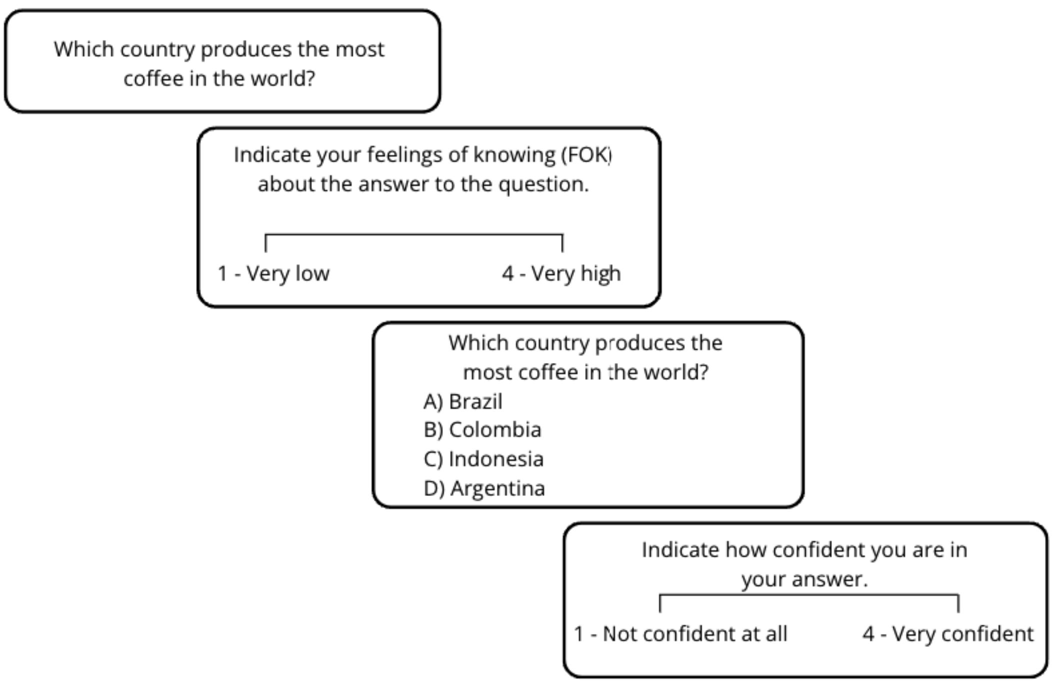
The Nonverbal Immediacy Scale – Self Report (NIS-SR; Richmond, et al., 2003) was translated to Turkish to examine the degree of participants' awareness of their gesture use. The scale consists of twenty-six statements (*i.e., I use my hands and arms while talking to others*), and the participants were asked to rate each statement from 1 (*never*) to 5 (*always*).

2.3 Procedure

After participants consented to participate in the study, they were sent to the experiment room. Participants were randomly assigned to visible listener or nonvisible listener conditions. The experimenter either sat face-to-face with the participant so that the listener was seen (visible condition), or the experimenter sat behind an opaque curtain so that only their voice could be heard by the participant (nonvisible condition). Participants were asked 40 general knowledge questions, such as "*How many eyes do bees have?*" and the questions were shown on a computer screen. Upon seeing each question, participants were first asked to provide feelings-of-knowing (FOK) judgments on a 4 point Likert scale (1: very low, 4: very high). This gave us a participant's prospective judgment on whether they have the feeling of knowing the answer to the open-ended question. Then, they were given the same question with 4AFC and were asked to vocalize their reasoning while pondering before answering. They were told to spend as much time as they needed on the questions. After each question, the participants were asked to provide a retrospective confidence judgment (RCJ) (i.e., *How confident are you in your answer?*) on a 4-point Likert scale (1: not confident at all, 4: very confident). This gave us a participant's metacognitive evaluation of first-order decisions (see Figure 1). There were two practice trials to help participants get used to the task. If they provided short or one-word answers, the experimenter gave feedback and asked them to vocalize their thoughts while thinking about the correct answer. Instructions did not involve anything regarding gesture use or speech disfluencies. The experimenter did not provide any feedback during the test trials and stayed entirely silent in both conditions while participants verbalized their reasoning in answering questions. Participants were informed that they would be video-recorded upon their consent. After they completed the task, they were asked to fill out the gesture awareness self-report measure (translated from Richmond et al., 2003). Materials, data, and analyses can be found at Open Science Framework (OSF) through the link:

https://osf.io/c64up/?view_only=1fa8d3b2a68f4d2293b68edf700378e5

Figure 1. *A sample trial from the experimental task*



2.4 Coding

All the word tokens produced were transcribed by using ELAN (version 6.6, Max Planck Institute for Psycholinguistics, 2023). We also coded hand gestures used and speech disfluencies while answering each question.

2.4.1 Speech disfluency coding

Disfluencies were originally categorized into four types by Maclay & Osgood (1959), namely filled pauses (i.e., "*uh*", "*um*"), silent (unfilled) pauses (i.e., *It just reminded me of… stars*), repeats (repetitions) (i.e., *I I don't really like it*.), and repairs (false starts) (i.e., *I was reading a newspa… magazine*.). Silent pauses were detected based on previous work, which suggests silence periods more than 200 ms can be classified as silent pauses, either within or between the sentences produced (Candea, 2000; but see Jong & Bosker, 2013 and Duez, 1982 for different minimum duration thresholds). For each question, the number of speech disfluencies were coded by type. In cases where one disfluency follows another (e.g., *I think it is blue.. um.. green*), we coded each disfluency (i.e., filled pause and repair) respectively. A native Turkish-speaking research assistant, who was naive to study hypotheses and experimental conditions coded all disfluencies and their types. A second coder selected 20% of the data randomly, which corresponds to 10 participants, and coded speech disfluencies. The ICC score of the two coders for the frequency of speech disfluencies was 96.7%. For each disfluency type, ICC scores were 79.1%, 95.5%, 79.4%, and 97.3% for filled pauses, repairs, repetitions, and silent pauses respectively. The discrepancies were resolved through discussion.

2.4.2 Gesture coding

For co-speech gestures, we used the categorization proposed by McNeill (1992), which includes iconic, metaphoric, deictic, and beat gestures. We also included palm-up gestures into the coding scheme as they were relevant to the question-answering paradigm used in the current study (Kendon, 2004). We classified gestures as representational and nonrepresentational. Representational gestures included iconic, metaphoric, and deictic gestures. For instance, when answering the question "*How many eyes do bees have?*", participants used iconic gestures, specifically making circular shapes with index and thumb fingers to represent eyes. Examples of metaphoric gestures were observed when the participants used their hands to represent abstract concepts such as mind illusion. Or, when participants used their index finger to point an empty space to locate eyes on a bee's face, such gestures were counted as deictic, referring to abstract representations of entities. Nonrepresentational gestures, on the other hand, included beat and palm-up gestures. A native Turkish-speaking research assistant blind to study hypotheses and experimental conditions coded all gestures and their types. Another researcher coded 20% of randomly selected data (10 participants). The ICC scores between two coders for gesture frequency was 97.3%. For the gesture type, ICC scores were 85.3% and 94.2% for representational and nonrepresentational gestures, respectively. The discrepancies were resolved via discussion.

2.5 Data analysis plan

There were 2000 trials (forty participants answered fifty questions) in total. We removed 14 trials (five participants did not produce verbal responses to one question, two participants produced no verbal response to two questions, and one participant did not verbally respond to five questions); remaining 1986 trials. For each of the trials, participants' accuracy (0: incorrect, 1: correct), feelings-of-knowing (FOK) judgments, and retrospective confidence judgments (RCJs) were considered. In addition, we calculated the total number of gestures and the number of gestures by type, as well as the total number of speech disfluencies and the number of disfluencies by type for each trial. Then, they were translated to percentages per 100 words to account for differences in words produced for each trial. For the main analyses, these percentages were added to the models at the trial level. For descriptives and correlation analyses, we averaged the percentage of both gestures and speech disfluencies across all trials for each participant.

We conducted analyses with R (version 4.4.0) (R Core Team, 2020). To test our hypothesis regarding the effect of communicational setting on the rate of speech disfluencies and co-speech gestures, we conducted a t-test. We used the package *ggpubr* (Kassambara & Kassambara, 2020). For the research questions related to participants' accuracy, metacognitive judgements, speech disfluencies, and gestures, we conducted generalized linear mixed effects and linear mixed effects analyses with the maximum likelihood ratio test. We used the packages *lme4* (Bates et al., 2014), and *lmerTest* to obtain p values (Kuznetsova et al., 2017). For post-hoc analyses, we used the package *emmeans* (Lenth et al., 2022). For visualization, we used the package *ggplot2* (Wickham et al., 2016). We included random intercepts in all models for the random factors of subjects and trials. Our outcome variables were accuracy, feelings-of-knowing (FOK) judgments, and retrospective confidence judgments (RCJs). We conducted generalized linear mixed models analysis to test participants' accuracy (0: incorrect, 1: correct), and we tested participants' FOK judgments and RCJs using linear mixed effects. For all analyses, we started with the simplest model, which includes only random intercepts (i.e., subjects and trials) without the inclusion of fixed effects. We then added our fixed effects (i.e., gestures and speech disfluencies), comparing the null model (i.e., the model including only random factors subjects and trials) with the full model (i.e., the model with both random factors and fixed effects) using *anova()* function. Both gestures and speech disfluencies were added to the models as standardized scores. We reported the full models for each analysis.

Results

Among 1986 trials, participants answered 844 trials (42.3%) correctly. In these trials, participants produced 72,857 words in total. On average, participants produced 36.57 words per question (*SD* = 15.29, *Range* = 10.87 - 83.35). Participants used at least one gesture in 1371 trials (69.03%) and had at least one speech disfluency in 1903 trials (95.82%).

Participants' accuracy scores correlated positively with their retrospective confidence judgments (RCJs) ($r = 0.411$, $p < .01$) but not with their feelings-of-knowing (FOK) judgments ($p > .05$). Similarly, there was no correlation between FOK judgments and RCJs ($p > .05$). We did not find a correlation between participants' overall gestures and their gesture awareness, either ($p > .05$). Thus, gesture awareness scores were not added to further analyses. Table 1 presents descriptive statistics and Table 2 shows correlations among all variables.

Table 1. *Descriptive statistics for the variables in the study*

	M	SD	Range
Age	21.16	1.46	18 - 24
Accuracy	16.9	4	10 - 27
Feelings-of-knowing (FOK)	2.12	0.33	1.22 - 2.9
Retrospective confidence judgment (RCJ)	2.08	0.39	1.25 - 2.85
Gesture awareness scale	91.88	14.31	54 - 116

Table 2. *Correlations between study variables*

	Accuracy	**FOK**	**RCJ**	**Overall gesture**	**Representational gesture**	**Nonrepresentational gesture**	**Overall disfluency**	**Gesture awareness**
Accuracy	—							
Feelings-of-knowing (FOK)	0.097	—						
Retrospective confidence judgment (RCJ)	0.411**	0.139	—					
Overall gesture	-0.08	-0.118	-0.161	—				
Representational gesture	-0.199	-0.043	-0.189	0.858***	—			
Nonrepresentational gesture	0.078	-0.161	-0.081	0.825***	0.417***	—		

Overall disfluency	-0.089	0.068	-0.015	0.060	0.096	0.000	—	
Gesture awareness	-0.190	0.011	-0.126	0.241	0.252	0.147	0.037	—

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

3.1 Are speech disfluencies and gestures related to accuracy and metacognitive judgments?

3.1.1 Accuracy

First, using generalized linear mixed models analysis, we asked whether gestures and speech disfluencies predicted participants' accuracy (0: incorrect, 1: correct). The model where fixed effects were included did not improve the model fit compared to the model with only random intercepts (i.e., subjects and trials) ($\chi 2(3)= 2.38$, $p = 0.496$). Results showed that neither gesture use ($\beta = 0.06$, $SE = 0.06$, $z = 1.01$, $p = .311$) nor speech disfluencies ($\beta = -0.06$, $SE = 0.05$, $z = -1.12$, $p = .260$) predicted accuracy. The interaction between gesture use and speech disfluencies predicting accuracy was not significant, either ($\beta = -0.02$, $SE = 0.03$, $z = -0.73$, $p = .464$). When question difficulty was added to the model as a fixed effect, it improved the model fit ($\chi 2(4)= 13.26$, $p < .001$). That is, question difficulty predicted accuracy, in which easier questions were answered more correctly than difficult questions. However, question difficulty did not interact with gesture use or speech disfluencies predicting accuracy (*ps* > .05).

3.1.2 Metacognitive judgments

Regarding our main questions, we examined whether metacognitive judgements (feelings-of-knowing (FOK) and retrospective confidence judgments (RCJs)) were related to gestures and speech disfluencies. The model predicting FOK judgments including the fixed effects did not contribute to improvement in the model fit ($\chi 2(3)= 0.42$, $p = 0.934$). Linear mixed effects model analysis revealed no significant effects of gesture ($b = -0.01$, $SE = 0.02$, $z = -0.35$, $p = .725$) or speech disfluency ($b = 0.00$, $SE = 0.02$, $z = 0.23$, $p = .811$) on FOK judgments. The interaction between gestures and speech disfluencies was not significant, either ($b = 0.00$, $SE = 0.01$, $z = 0.58$, $p = .560$). When question difficulty was added to the model as a fixed effect, the model showed no improvement ($\chi 2(4)= 1.63$, $p = .803$).

For the retrospective confidence judgements, the model with the fixed effects (i.e., overall gestures and speech disfluencies) improved the model fit ($\chi 2(3)= 44.93$, $p < .001$). Linear mixed effects analysis showed a significant effect of overall gesture use ($b = 0.07$, $SE = 0.02$, $z = 3.09$, $p = .002$) and speech disfluency ($b = -0.08$, $SE = 0.02$, $z = -4.10$, $p < .001$) on RCJs (see Figure 2). Those participants who used more gestures had higher RCJs. Conversely, speech disfluencies predicted confidence judgments negatively, such that participants who produced more speech disfluencies provided lower retrospective confidence judgements. Additionally, the interaction between gestures and speech disfluencies was also significant ($b = 0.04$, $SE = 0.01$, $z = 3.26$, $p = .001$) (see Figure 3). That is, the increased use of gestures was more strongly related to confidence judgments when disfluencies were also higher (see Table 3 for fixed effect estimates). When we further included question difficulty to the model, it did not show any improvement in the model ($\chi 2(4)= 0.35$, $p = .986$). Thus, question difficulty did not add variability to RCJs, and we observed the effects of speech disfluencies and gestures on RCJs regardless of varying question difficulties. To evaluate whether our sample size was sufficient to detect the observed effects, we have conducted simulation based post-hoc power analysis in R using *simr()* package with 1,000 simulations for each fixed effect with alpha set to 0.05 (Kumle et al., 2021). The estimated power to detect the main effect of gesture use with our sample size was 87.9% for a small effect size (0.07). The main effect of speech disfluency showed 99% power with the effect size of -0.08, and their interaction achieved 90.1% power with the effect size 0.04. These results indicate that the model was adequately powered to detect small-to-moderate effect sizes.

Further, we conducted exploratory analyses to understand what type of gestures (i.e., representational and nonrepresentational) were related to retrospective confidence judgments (see Table 3). The model including representational gestures (i.e., deictic, iconic, and metaphoric) showed a main effect of these gestures on confidence judgments ($b = 0.08$, $SE = 0.02$, $z = 3.38$, $p < .001$). That is, participants who used more representational gestures provided higher retrospective confidence judgements. However, there was no interaction between representational gestures and speech disfluencies on predicting confidence judgments ($b = 0.02$, $SE = 0.01$, $z = 1.84$, $p = .065$)[1]. Unlike representational gestures, the model with nonrepresentational gestures (i.e., beat and palm-up) showed no main effect of nonrepresentational gestures on confidence judgments ($b = 0.02$, $SE = 0.02$, $z = 1.20$, $p = .228$). However, the interaction between nonrepresentational gestures and speech disfluencies was significant ($b = 0.03$, $SE = 0.01$, $z = 2.95$, $p = .003$). Post hoc analysis showed that when participants used nonrepresentational gestures along with speech disfluencies, they provided higher confidence judgments.

Figure 2. *Predicted confidence judgments by A) speech disfluencies and B) hand gestures.*

A) B)

[1] When deictic gestures were removed from the category of representational gestures, results did not change. The main effect of representational gestures was significant ($b = 0.07$, $SE = 0.02$, $z = 3.30$, $p < .001$), and the interaction remained nonsignificant ($b = 0.02$, $SE = 0.01$, $z = 1.63$, $p = .103$).

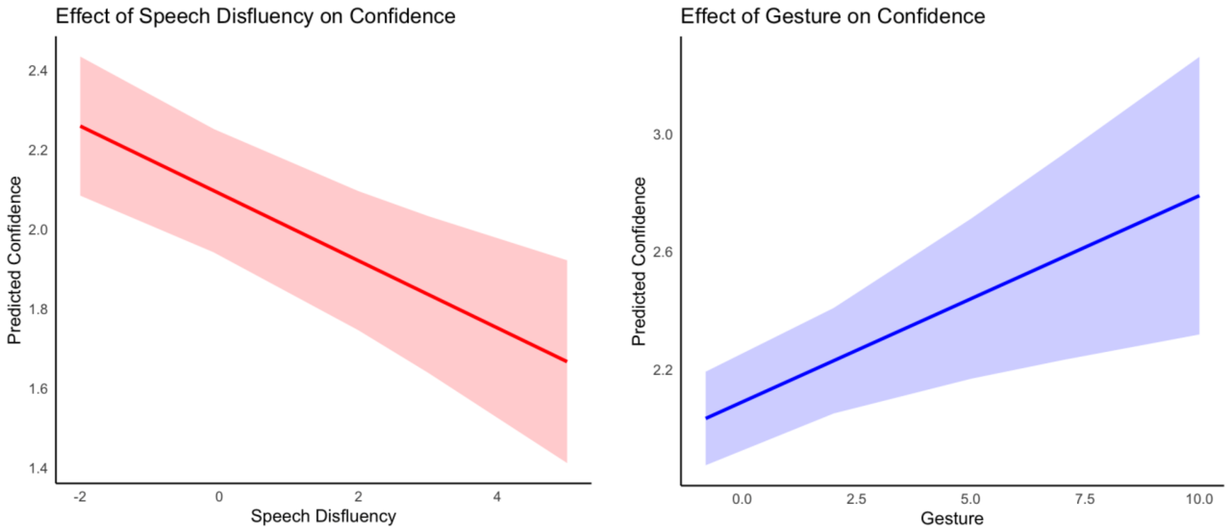


Figure 3. *Predicted retrospective confidence ratings as a function of gesture use and speech disfluencies. When speech disfluency increases 1 SD, the relation between overall gesture use and retrospective confidence judgments (RCJs) becomes stronger.*

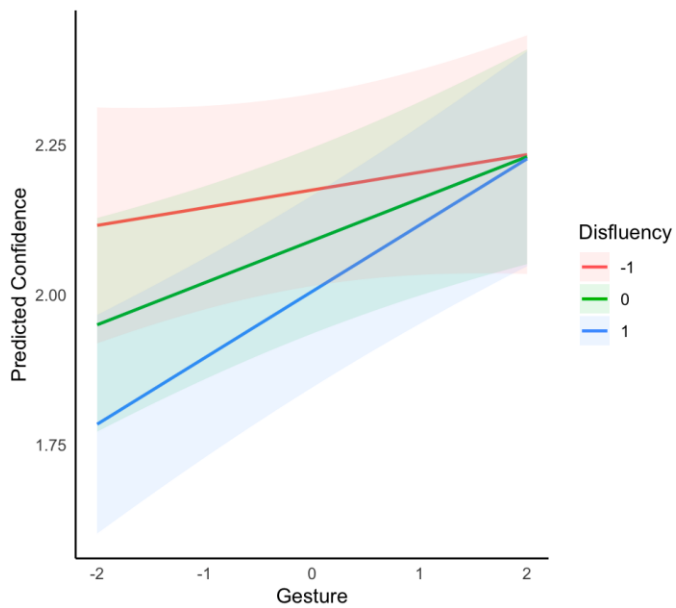


Table 3. *Fixed effect estimates from linear mixed-effects models for retrospective confidence judgments (RCJs)*

	Model 1 with overall gestures			Model 2 with representational gestures			Model 3 with nonrepresentational gestures		
	b (SE)	z	p	*b* (SE)	z	p	*b* (SE)	z	p
Fixed effects									
(Intercept)	2.08 (0.07)	26.44	**< .001**	2.09 (0.07)	26.38	**< .001**	2.09 (0.07)	26.81	**< .001**
Gesture	0.07 (0.02)	3.09	**0.002**	0.08 (0.02)	3.38	**< .001**	0.02 (0.02)	1.20	0.228
Disfluency	-0.09 (0.02)	-4.1	**< .001**	-0.08 (0.02)	-4.07	**< .001**	-0.07 (0.02)	-3.69	**< .001**
Gesture:Disfluency	0.04 (0.01)	3.26	**0.001**	0.02 (0.01)	1.84	0.065	0.03 (0.01)	2.95	**0.003**

Notes. Bold typeface indicates a significant effect.

3.2 Does the communicational setting change the rate of speech disfluencies and gestures?

Results revealed no effect of listener visibility on gesture production or speech disfluency. Independent samples t-test revealed no significant difference in overall gesture use or different gesture types ($t(48) = -1.03$, $p = 0.309$ for overall gesture use, $t(48) = -1.13$, $p = 0.266$ for representational gesture use, and $t(48) = -0.59$, $p = 0.553$ for nonrepresentational gesture use). Welch's test also indicated no difference in overall speech disfluency rates $t(41.3) = -0.52$, $p = 0.300$ nor different types (*ps* > .05). As such, we did not conduct additional analyses including the listener visibility effect in the main models reported above. See Table 4 for gestures by type and Table 5 for disfluencies by type per 100 words with the change in communicational setting.

Table 4. *Gesture rates on average per 100 words*

	Visible listener (*n* = 25)			Nonvisible listener (*n* = 25)		
	M	SD	Range	M	SD	Range
Representational gestures	3.09	2.29	0.1 - 8.99	4.38	5.25	0.08 - 23.4
Nonrepresentational gestures	5.52	3.22	0.67 - 13.5	6.15	4.16	0.44 - 15.8
Overall gestures	8.61	4.91	0.81 - 20.1	10.5	7.81	0.92 - 29.7

Table 5. *Speech disfluency rates on average per 100 words*

	Visible listener (*n* = 25)			Nonvisible listener (*n* = 25)		
	M	SD	Range	M	SD	Range
Filled pause	2.16	1.91	0.28 - 7.54	3.13	2.43	0.16 - 9.18
Repair	0.46	0.34	0 - 1.21	0.42	0.32	0 - 1.52
Repetition	0.57	0.52	0 - 1.8	0.42	0.31	0 - 1
Silent pause	13.9	3.82	6.9 - 21.7	13.9	5.55	3.86 - 24.7
Overall disfluency	17.06	4.26	9.87 - 25.49	17.88	6.53	8.21 - 34.04

Discussion

The current study investigated whether hand gestures and speech disfluencies were related to metacognitive judgments, and whether the rate of gestures and speech disfluencies changed as a function of the communicational setting. We asked participants general knowledge questions and they provided their answers aloud, along with their prospective (FOK) and retrospective (RCJs) metacognitive judgments. We found that speech disfluencies and gestures were related only to retrospective confidence judgments (RCJs) but in different directions. Participants provided lower RCJ ratings when they produced more speech disfluencies. In contrast, they provided higher RCJs when they used more gestures. More importantly, speech disfluencies modulated the effect of co-speech gestures on RCJs. The impact of gestures on confidence increased with higher rates of speech disfluencies. Listener visibility did not have a significant effect on the production of speech disfluencies or gestures. These findings, for the first time, suggest that multimodal language related mechanisms can be important cues in the formation of metacognitive judgments.

4.1 Linking speech disfluencies and gestures to accuracy and metacognitive judgments

4.1.1 Accuracy

With increased question difficulty, accuracy diminished in the current study. However, neither speech disfluencies nor gesture production changed with varying degrees of accuracy. Earlier studies show that people use more gestures when task difficulty increases, as a way of reducing their cognitive load in different tasks, such as spatial reasoning (Çapan et al., 2024) and picture description (Melinger & Kita, 2007). However, our general-question answering paradigm did not reveal such an association. The fact that question difficulty affected only performance and not metacognitive judgments, disfluency or gestures, further suggests that gestures and disfluencies might be metacognitive products working retrospectively. Once the participant starts tackling a question, consequences of that process inform metacognitive judgements and call for adjustments in cognitive operations.

4.1.2 Speech disfluencies

In line with our hypothesis, results showed that participants who had more speech disfluencies provided lower retrospective confidence ratings. Unlike previous work, the feelings-of-knowing judgments were not related to disfluencies in speech in the current study (Smith & Clark, 1993; Swerts & Krahmer, 2005). Studies show that not all metacognitive judgements (i.e. judgements-of-learning, feelings-of-knowing, retrospective confidence judgements) are subserved by the same processes (Chua & Solinger, 2015; Fleming et al., 2016; Mazancieux et al., 2020). As both judgments-of-learning (JOL) and feelings-of-knowing (FOK) are commonly given prospectively and confidence judgments are given retrospectively, the cues affecting these metacognitive judgments can also differ. Supporting this, we did not find a correlation between FOKs (i.e. prospective metacognitive judgment) and RCJs (i.e., retrospective metacognitive judgment). It is likely that the cues elicited during the task at hand (e.g. fluency) might directly contribute to the construction of metacognitive judgments. Supporting this, researchers suggest that when providing a retrospective confidence rating, rather than judgments-of-learning (JOL), people predominantly rely on cues elicited during online processing (Authors, in prep; Mueller & Dunlosky, 2017). Although JOLs can be affected by both prior beliefs and fluency during learning, retrospective confidence judgments would also benefit from the cues elicited during the task. The current study suggests that speech disfluencies act as such a cue for the metacognitive system, independent of first-order performance.

4.1.3 Gestures

Our results regarding the relation between gestures, first-order performance, feelings-of-knowing (FOK), and retrospective metacognitive judgments showed that gestures were not related to task accuracy or feelings-of-knowing (FOK) judgments. However, we showed that gesture use was related to metacognitive judgments when given retrospectively. Similar to speech disfluencies, gestures might function as cues during online processing, in our case creating an illusory effect on the metacognitive system. However, contrary to our hypothesis that gesture use would reduce RCJs, hand gestures increased RCJs. That is, gestures were associated with an increase in metacognitive evaluations of one's past performance. Higher gesture production did not improve task performance but increased RCJs, creating a metacognitive illusion.

How exactly do gestures lead to metacognitive illusions, associated with an increase in people's confidence while not related to the performance? Previous research provides evidence that embodied actions might modulate metacognitive processes without affecting accuracy (Alban & Kelley, 2013; Palser et al., 2018). The direction of the link between embodied cues and metacognitive judgments may depend on whether these cues are used as part of top-down or bottom-up processes, reflecting different investments of effort (Koriat et al., 2006). We would have expected a negative relationship between gestures and metacognitive judgments if gestures had been produced as a result of inherent difficulty associated with the task, acting as bottom-up cues. However, our results showed a positive link between gestures and metacognitive judgments, favoring the hypothesis that gestures were produced as a result of metacognitive strategic allocation. In this case, gestures might have been used purposefully, in a top-down fashion. However, it is also possible that both top-down and bottom-up processes are simultaneously at play (Koriat et al., 2006; Koriat & Nussinson, 2009). People may have produced gestures based on their subjective confidence as a top-down cue due to the judged difficulty of the task, but also the feedback from gesture use might inform metacognitive evaluations. In that sense, gestures might be used in a dynamic fashion. Gestures might have also created a sense of ease or fluency on the metacognitive system, resulting in increased confidence. Similar to what we have found, a recent study also provided evidence that those who were encouraged to use gestures during spatial reasoning provided higher retrospective confidence ratings (Çapan et al., 2024). It is possible that, although gestures were produced spontaneously in the current study, reflecting a top-down process, they may have also served as a bottom-up influence on the subjective estimates of confidence. These results together suggest that cues elicited through embodied actions (i.e., gestures) might carry metacognitive information inflicting both data-driven and goal-driven influence on subsequent higher-order decisions. In sum, the current study adds to the literature that gestures not only unfold thoughts, but can also unfold thoughts about thoughts, although in a way that creates illusions.

It could also be argued that off-line metacognitive judgments might influence gesture production. That is, the higher one's feelings-of-knowing, the more people gesture. However, our additional analysis on whether feelings-of-knowing judgments were related to gesture use showed no significant results, suggesting that gestures do not reflect higher feelings-of-knowing (see Supplementary Materials Table S1). Although in our study we did not experimentally manipulate gesture use and cannot claim any causal relationship between gestures and metacognitive processes, our analysis partially ruled out the possibility that gesture use was affected by one's prospective metacognitive judgments.

4.1.4 Speech disfluency and gesture interaction

Our results have shown that although speech disfluencies suppressed confidence judgements in general, this effect was dulled with the use of gestures in a given trial. It is possible that speech disfluencies create a sense of difficulty, which then are mitigated by the use of gestures. Earlier studies provide evidence that when gesture use is restricted, people produce more disfluencies compared to when they freely use gestures (Finlayson et al., 2003; Morsella & Krauss, 2004; Rauscher et al., 1996). This finding suggests that the presence of gestures might act as an alternative channel to convey information through speech, helping the speaker express their thoughts more clearly. When an error in speech occurs, gestures are also used to help resolve these temporary speech problems. This is in line with earlier studies showing that gestures accompany disfluent speech more (Finlayson et al., 2003; Arslan et al., 2022; 2024; but see Akhavan et al., 2016; Avcı et al., 2022; Cravotta et al., 2021; Hoetjes et al., 2014). Then, it would be the case that when informing the metacognitive system, gestures afford a sense of ease counteracting the effect of speech disfluencies. The activation of the motoric system when a speech error is detected might function as a compensatory mechanism, particularly in an intact and developed language network, informing the metacognitive system, and increasing confidence. Once people produce gestures and speech disfluencies, these self-related cues may together inform the monitoring system, making the monitoring process retrospective in nature. In other words, it is possible that rather than the metacognitive system triggering language related cues, (e.g. a difficult question calling for gestures for aid), the consequences of first-order cognitive processes (i.e. gestures and speech disfluencies) inform the metacognitive system through specific cues. However, although confidence judgments were temporally made after the language related cues are elicited, this argument remains untested due to the correlational nature of the current study. Future studies could experimentally manipulate gesture use and investigate whether speech disfluencies and metacognitive judgments change accordingly.

Overall, our results suggest a complex relationship among gesture, speech disfluency, and metacognitive judgments. On the one hand, speech disfluencies may be used as bottom-up cues and the feedback retrieved from speech disfluencies may have informed subsequent metacognitive evaluations as reflected in lower estimates of confidence. On the other hand, gestures might be produced as a function or as a result of monitoring, used as both bottom-up and top-down cues (Koriat et al., 2006). In that sense, although gestures are produced intentionally as a top-down fashion, the sensory feedback retrieved from gesture use might have informed confidence judgments as a bottom-up cue creating the sense of competence, explaining why gestures are related to higher confidence. Beyond that, the buffering effect of gestures on speech disfluencies dampening its negative relationship with confidence also suggest that multiple cues are at play informing metacognitive judgments. Our results are in line with Koriat et al. (2006), suggesting that the assessment and weighing of different available cues might require the involvement of both top-down and bottom-up processes on metacognitive evaluations. Therefore, the dynamic interplay between speech disfluencies and gestures might be a great avenue to examine the link between embodied cognition and metacognition considering the different mental efforts associated with each and their interaction.

Existing research focusing on gestures and metacognition (Çapan et al., 2024) takes into account representational gestures (i.e., iconic and metaphoric gestures) which are related to the semantic content of speech, neglecting nonrepresentational gestures as they are claimed to be unrelated to speech content. However, researchers also argue that nonrepresentational gestures might be used with internal functions in speech beyond their function as visual focus markers (Alibali et al., 2001; Ferrari & Hagoort, 2025; Prieto et al., 2018). For instance, Ferrari & Hagoort (2025) showed that people make inferences regarding the speaker's confidence when beat gestures (i.e., the rhythmic hand movements) are used along with speech. These rhythmic movements during speech that are related to the speech prosody might be used with speaker-internal functions, thus may also inform metacognitive judgments. In other words, whether hand gestures represent meaning or not, the activation of the motor system can affect both speech disfluencies and metacognitive judgments. In our additional analysis teasing apart the effects of different types of gestures, we found that only representational gestures, but not nonrepresentational gestures were related to metacognitive judgments. Rather than beat gestures or gestures with pragmatic functions (i.e., palm up), those that represent meaning with hands were responsible for creating metacognitive illusions. Although our results do not show a main effect of nonrepresentational gestures on metacognitive judgments, the model showed an interaction between nonrepresentational gestures and speech disfluencies on retrospective confidence judgments; those individuals who used nonrepresentational gestures along with disfluencies also had higher retrospective confidence scores. This was not seen for representational gestures. Thus, although nonrepresentational gestures were not related to metacognitive judgments directly, they were still effective in mitigating the effect of speech disfluencies on metacognitive judgments.

4.2 The effect of the visibility of listener on speech disfluencies and gestures

4.2.1 On speech disfluencies

Contrary to earlier studies (Alibali et al., 2001; Kasl & Mahl, 1965; Oviatt, 1995; Rimé, 1982) and our prediction, results did not show an effect of the visibility of the listener on speech disfluency rates. Our null findings regarding the effect of listener visibility on the production of speech disfluencies could be attributed to the visibility manipulation in the current study. The experimenter in the visible condition did not provide any feedback to the participants in any modality (i.e., speech, gesture, nodding, etc.). Thus, although the speaker could provide information in other modalities (i.e., gestures, eye gaze or nodding) when the listener was visible, there was no back-channel response from the listener to coordinate conversation. It is possible that the visibility manipulation in our experiment was not regarded as participants having a conversation, rather it might be regarded more as a monologue. It would be that in dialogues, conversational partners produce multimodal cues at a higher rate, enabling the speaker to align their speech better to keep up with the conversation. Future designs should consider including an interactive listener to test the effect of listener visibility on the rate of speech related cues to understand the mechanism in which these cues change as a function of the communicational setting.

4.2.2 On gestures

Similar to speech disfluencies, our results did not show an effect of listener visibility on overall, representational or non-representational gesture use. The potential reason would be that the mere presence of a listener might be seen as an unresponsive or less attentive addressee, leading to consider the conversation as a monologue rather than a dialogue. There is evidence showing that unresponsive addressees reduce the overall gesture rate compared to a responsive addressee (Beattie & Aboudan, 1994). Researchers also argue that the effect of dialogue on gesture production might be more robust than the effect of listener visibility (Bavelas et al., 2008; see Bavelas & Healing, 2013 for a review). Besides, earlier studies that show the visibility effect on the gesture production rates mostly use confederates as the addressees, creating a quasi-dialogue which may not be an ideal experimental control as opposed to a free dialogue (Bavelas & Healing, 2013). Further, it is also possible that the absence of visibility effect on different gesture types in the current study is related to the nature of the task. Unlike spatial or motoric tasks, which often elicit self-oriented gestures, the general-knowledge questions used in the current study are more likely to elicit representational gestures with a communicative purpose. In such cases, gestures may be driven more by the presence of a conversational partner than by the visibility of the listener. Thus, rather than the mere presence of a listener, multimodal cues retrieved from the interactive partner could make one engage in richer ways to inform the other than solely sharing information using only spoken language. Overall, our null findings could be attributed to the design in the current study, thus, requiring future research to investigate whether interactive designs resembling a dialogue might influence the rate of gestures produced.

Additionally, studies focusing on the communicative functions of gestures, which manipulate listener visibility or dialogue-monologue conditions, disregard the self-oriented functions of gestures in most cases (Alibali et al., 2001; Bavelas et al., 2002; 2008; Krauss et al., 1995). Although communicative and self-oriented purposes of gesture use are not mutually exclusive, in our study, the cognitive functions of gestures prevailed. If people keep producing gestures even when they cannot see the listener, then gestures are not merely produced to be seen, rather, they might be produced with self-oriented functions servicing cognitive purposes.

With regard to individual differences in gesture production, self-report assessment on gesture use awareness did not reveal any association with overall gesture use. It is possible that self-report measures of gesture use might not reliably assess people's awareness of their and others' gesture use, necessitating investigation of gesture use in different tasks. Further studies might use diverse tasks to understand whether people use gestures at a similar rate in different tasks, suggesting individual differences in gesture production, and how they are related to contextual differences such as the visibility of the listener.

4.3 Limitations and future directions

Together with speech disfluencies and gestures, there can be other paralinguistic cues that might be related to one's metacognitive judgments such as shrugs, speech prosody, eye gaze, head and brow movements, or facial expressions (Swerts & Krahmer, 2005), which were not investigated in the current study. These cues can also be informative of the person's internal states, and future studies might also investigate how these cues can contribute to metacognitive judgments.

Although our study does not provide a causal explanation regarding the relation between gesture-speech disfluency co-occurrences in a given trial and metacognitive judgments, it points to a causal mechanism. Future work would manipulate gesture use and investigate how gesture restriction or encouragement affect disfluency rates and how they are together related to metacognitive judgments. Further research can also investigate which gesture types are related to speech disfluencies and metacognitive judgments in a question answering paradigm by manipulating or encouraging representational and/or nonrepresentational gesture use.

In the current study, we examined gestures and speech disfluencies in the same trial, not their temporal alignment. That is, in a given trial, one might produce both hand gestures and speech disfluencies at different time points. As this is a contextual paradigm which does not aim to reveal any relation regarding lexical retrieval at the word level, we did not conduct analyses regarding when and how gestures and speech disfluencies align temporally. Thus, our results must be evaluated cautiously while making generalizations on gesture speech disfluency co-occurrences. A future direction could be to examine the temporal alignment of gestures and speech disfluencies, and how they are related to one's confidence judgments from both gesture and speech disfluency perspectives (i.e., gestures accompanying speech disfluencies and speech disfluencies accompanying gestures) in a different paradigm (see Arslan et al., 2022).

Although our results provide insight on metacognitive judgments, how gesture use and speech disfluency are related to *metacognitive efficiency* is an understudied area. In the current study, we did not calculate individuals' m-ratio scores, which measures metacognitive efficiency. Future research might investigate whether those who have higher metacognitive efficiency (i.e., providing confidence rating in line with the performance) on the task are also more likely to use gestures and speech disfluencies.

Furthermore, our sample was not homogenous, including 32 females (18 males). Previous research has found gender differences in metacognitive processes (Lemieux et al., 2019; Rivers et al., 2021). Thus, further studies may consider recruiting equal size of both genders to investigate whether there are gender differences in informing the metacognitive system with multimodal cues.

To establish a robust framework on first-order and higher-order processes (i.e., metacognitive monitoring and control), and how they are related to gestures and speech disfluencies, the neural underpinnings of metacognitive processes can also be studied. Brain and or scalp imaging methods that provide high temporal resolution (i.e., EEG) might be appropriate to establish a framework for the interplay between metacognitive monitoring and control; the ways in which one mechanism informs the other, their relation to first-order performance, and behavioral outcomes.

5. Conclusion

In the current study, we examined whether hand gestures and speech disfluencies predicted metacognitive judgments, by acting as a metacognitive cue, and whether the visibility of the listener changed the rate of speech disfluencies and gestures produced. This is the first study investigating both the contextual factors associated with gesture and speech disfluency production, and how these processes interacted with higher-order cognitive mechanisms. Our findings indicated that both speech disfluencies and hand gestures act as metacognitive cues, both uniquely and interactively. This suggests that language related cues can inform metacognitive processes even when the first-order performance remains unrelated, thus yielding metacognitive illusions.

References

Akhavan, N., Göksun, T., & Nozari, N. (2016). Disfluency production in speech and gesture. *Cognitive Science Society*.

Alban, M. W., & Kelley, C. M. (2013). Embodiment meets metamemory: weight as a cue for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1628. 10.1037/a0032420

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen. *Journal of Memory and Language*, *44*(2), 169–188. https://doi.org/10.1006/jmla.2000.2752

Arslan, B., & Göksun, T. (2022). Aging, Gesture Production, and Disfluency in Speech: A Comparison of Younger and Older Adults. *Cognitive Science*, *46*(2), e13098. https://doi.org/10.1111/cogs.13098

Arslan, B., Avcı, C., Yılmaztekin, A., & Göksun, T. (2024). Do bilingual adults gesture when they are disfluent?: Understanding gesture-speech interaction across first and second languages. *Language, Cognition and Neuroscience*, *39*(5), 571–583. https://doi.org/10.1080/23273798.2024.2345306

Avcı, C., Arslan, B., & Göksun, T. (2022). Gesture and Speech Disfluency in Narrative Context: Disfluency Rates in Spontaneous, Restricted, and Encouraged Gesture Conditions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/6xg8p5rd

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4* (arXiv:1406.5823). arXiv. https://doi.org/10.48550/arXiv.1406.5823

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520. https://doi.org/10.1016/j.jml.2007.02.004

Bavelas, J., Gerwing, J., & Healing, S. (2014a). Effect of dialogue on demonstrations: Direct quotations, facial portrayals, hand gestures, and figurative references. *Discourse Processes*, *51*(8), 619-655. https://doi.org/10.1080/0163853X.2014.883730

Bavelas, J., Gerwing, J., & Healing, S. (2014b). Hand and facial gestures in conversational interaction.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, *13*(1), 63-92. https://doi.org/10.1075/gest.13.1.03bav

Bavelas, J., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture*, *2*(1), 1–17. https://doi.org/10.1075/gest.2.1.02bav

Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, *99*(3-4), 239-272. doi.org/10.1515/semi.1994.99.3-4.239

Beattie, G. W., & Butterworth, B. L. (1979). Contextual Probability and Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech. *Language and Speech*, *22*(3), 201–211. https://doi.org/10.1177/002383097902200301

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, *44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Brennan, S. E. (2002). Visual Co-Presence, Coordination Signals, and Partner Effects in Spontaneous Spoken Discourse. 認知科学, *9*(1), 7–25. https://doi.org/10.11225/jcss.9.7

Brennan, S. E., & Williams, M. (1995). The Feeling of Another′s Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language*, *34*(3), 383–398. https://doi.org/10.1006/jmla.1995.1017

Canarslan, F., & Chu, M. (2024). Individual differences in representational gesture production are associated with cognitive and empathy skills. *Quarterly Journal of Experimental Psychology*, 17470218241245831. https://doi.org/10.1177/17470218241245831

Candea, M. (2000). Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation» en français oral spontané. Étude sur un corpus de récit en classe de français. [Unpublished PhD Thesis]. Université Sorbonne Nouvelle – Paris III.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

Chua, E. F., & Solinger, L. A. (2015). Building metamemorial knowledge over time: Insights from eye tracking about the bases of feeling-of-knowing and confidence judgments. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01206

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, *28*(2), 276–279. https://doi.org/10.1037/h0035792

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign. *Frontiers in Communication, 3*. https://doi.org/10.3389/fcomm.2018.00023

Corley, M., & Stewart, O. W. (2008). Hesitation Disfluencies in Spontaneous Speech: The Meaning of um. *Language and Linguistics Compass*, *2*(4), 589–602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

Corley, M., Corley, M., Hartsuiker, R. J., & Hartsuiker, R. (2003). *Hesitation in speech can. . . Um. . . Help a listener understand*.

Cravotta, A., Prieto, P., & Busà, M. G. (2021). Exploring the effects of restraining the use of gestures on narrative speech. *Speech Communication*, *135*, 25–36. https://doi.org/10.1016/j.specom.2021.09.005

Çapan, D., Furman, R., Göksun, T., & Eskenazi, T. (2024). Hands of confidence: When gestures increase confidence in spatial problem-solving. *Quarterly Journal of Experimental Psychology*, 174702182311642. https://doi.org/10.1177/17470218231164270

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the ( ): Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Double, K. S., & Birney, D. P. (2017). Are you sure about that? Eliciting confidence ratings may influence performance on Raven's progressive matrices. *Thinking & Reasoning*, *23*(2), 190–206. https://doi.org/10.1080/13546783.2017.1289121

Duez, D. (1982). Silent and Non-Silent Pauses in Three Speech Styles. *Language and Speech*, *25*(1), 11–28. https://doi.org/10.1177/002383098202500102

ELAN (Version 6.6) [Computer software]. (2023). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Ferrari, A., & Hagoort, P. (2025). Beat gestures and prosodic prominence interactively influence language comprehension. *Cognition*, *256*, 106049. https://doi.org/10.1016/j.cognition.2024.106049

Ferré, G. (2010). *Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French*.

Finlayson, S., Forrest, V., Lickley, R., & Beck, J. M. (2003). *Effects of the restriction of hand gestures on disfluency*. https://eresearch.qmu.ac.uk/handle/20.500.12289/2180

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological science*, *26*(1), 89-98. https://doi.org/10.1177/0956797614557697

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, *2016*(1), niw018. https://doi.org/10.1093/nc/niw018

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, *29*(9), 1083-1096. https://doi.org/10.1080/01690965.2013.832785

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2213–2223. https://doi.org/10.1098/rstb.2012.0123

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. https://doi.org/10.1111/brv.12535

Galati, A., Weisberg, S. M., Newcombe, N. S., & Avraamides, M. N. (2018). When gestures show us the way: Co-thought gestures selectively facilitate navigation and spatial memory. *Spatial Cognition & Computation*, *18*(1), 1–30. https://doi.org/10.1080/13875868.2017.1332064

Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (2013). Naming and gesturing spatial relations: Evidence from focal brain-injured individuals. *Neuropsychologia*, *51*(8), 1518-1527.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429. https://doi.org/10.1016/S1364-6613(99)01397-2

Hoetjes, M., Krahmer, E., & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, *57*, 257–267. https://doi.org/10.1016/j.specom.2013.06.007

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, *7*(1), 73–95. https://doi.org/10.1075/gest.7.1.05hos

Jong, N. H. D., & Bosker, H. R. (2013). *Choosing a threshold for silent pauses to measure second language fluency*.

Kasl, S. V., & Mahl, G. F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, *1*(5), 425–433. https://doi.org/10.1037/h0021918

Kassambara, A., & Kassambara, M. A. (2020). Package 'ggpubr'. *R package version 0.1*, *6*(0).

Kendon, A. (1988). How gestures can become like words. In *Cross-cultural perspectives in nonverbal communication* (pp. 131–141). Hogrefe & Huber Publishers.

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

Kimble, C. E., & Seidel, S. D. (1991). Vocal signs of confidence. *Journal of Nonverbal Behavior*, *15*(2), 99–105. https://doi.org/10.1007/BF00998265

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological review*, *124*(3), 245. https://doi.org/10.1037/rev0000059

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Koriat, A., & Nussinson, R. (2009). Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1338–1343. https://doi.org/10.1037/a0016374

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The Communicative Value of Conversational Hand Gesture. *Journal of Experimental Social Psychology*, *31*(6), 533–552. https://doi.org/10.1006/jesp.1995.1024

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior research methods*, *53*(6), 2528-2543. https://doi.org/10.3758/s13428-021-01546-0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lemieux, C. L., Collin, C. A., & Watier, N. N. (2019). Gender differences in metacognitive judgments and performance on a goal-directed wayfinding task. *Journal of Cognitive Psychology*, *31*(4), 453-466. https://doi.org/10.1080/20445911.2019.1625905

Lenth Russell, V., Paul, B., Maxime, H., Maarten, J., Jonathon, L., Fernando, M., ... & Henrik, S. (2022). emmeans: Estimated marginal means, aka least-squares means. *R package version 1.7*, *5*.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Lickiss, K. P., & Wellens, A. R. (1978). Effects of Visual Accessibility and Hand Restraint on Fluency of Gesticulator and Effectiveness of Message. *Perceptual and Motor Skills*, *46*(3), 925–926. https://doi.org/10.2466/pms.1978.46.3.925

Maclay, H., & Osgood, C. E. (1959). Hesitation Phenomena in Spontaneous English Speech. WORD, *15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

Marstaller, L., & Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin & Review*, *20*(3), 496–500. https://doi.org/10.3758/s13423-012-0365-0

Mazancieux, A., Dinze, C., Souchay, C., & Moulin, C. J. A. (2020). Metacognitive domain specificity in feeling-of-knowing but not retrospective confidence. *Neuroscience of Consciousness*, *2020*(1), niaa001. https://doi.org/10.1093/nc/niaa001

Mazor, M., Dijkstra, N., & Fleming, S. M. (2022). Dissociating the Neural Correlates of Subjective Visibility from Those of Decision Confidence. *Journal of Neuroscience*, *42*(12), 2562–2569. https://doi.org/10.1523/JNEUROSCI.1220-21.2022

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

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*(4), 473–500. https://doi.org/10.1080/01690960600696916

Merlo, S., & Mansur, L. L. (2004). Descriptive discourse: Topic familiarity and disfluencies. *Journal of Communication Disorders*, *37*(6), 489–503. https://doi.org/10.1016/j.jcomdis.2004.03.002

Morson, S. M., Moulin, C. J., & Souchay, C. (2015). Selective deficits in episodic feeling of knowing in ageing: A novel use of the general knowledge task. *Acta Psychologica*, *157*, 85-92.

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, *93*, 245-258.

Nelson, T. O. (1990). Metamemory: A Theoretical Framework and New Findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. *Errors in Linguistic Performance : Slips of the Tongue, Ear, Pen and Hand / Ed. by Victoria A. Fromkin*, 87–95.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. https://doi.org/10.1016/j.tics.2008.02.014

Oviatt, S. (1995). Predicting spoken disfluencies during human–computer interaction. *Computer Speech & Language*, *9*(1), 19–35. https://doi.org/10.1006/csla.1995.0002

Özer, D., & Göksun, T. (2020). Gesture Use and Processing: A Review on Individual Differences in Cognitive Resources. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.573555

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, *28*, 151–160. https://doi.org/10.1016/j.concog.2014.06.007

Palser, E. R., Fotopoulou, A., & Kilner, J. M. (2018). Altering movement parameters disrupts metacognitive accuracy. *Consciousness and cognition*, *57*, 33-40.

Patel, D., Fleming, S. M., & Kilner, J. M. (2012). Inferring subjective states through the observation of actions. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1748), 4853-4860. https://doi.org/10.1098/rspb.2012.1847

Pine, K. J., Gurney, D. J., & Fletcher, B. (2010). The semantic specificity hypothesis: When gestures do not depend upon the presence of a listener. *Journal of Nonverbal Behavior*, *34*, 169-178. https://doi.org/10.1007/s10919-010-0089-7

Prieto Vives, P., Cravotta, A., Kushch, O., Rohrer, P. L., & Vilà-Giménez, I. (2018). *Deconstructing beat gestures: A labelling proposal*. https://doi.org/10.21437/SpeechProsody.2018-41

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological science*, *7*(4), 226-231. https://doi.org/10.1111/j.1467-9280.1996.tb00364.x

Richmond, V. P., McCroskey, J. C., & Johnson, A. D. (2003). Development of the nonverbal immediacy scale (NIS): Measures of self-and other-perceived nonverbal immediacy. *Communication Quarterly*, *51*(4), 504–517. https://doi.org/10.1080/01463370309370170

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology*, *12*(2), 113–129. https://doi.org/10.1002/ejsp.2420120201

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, *36*, 61–74. https://doi.org/10.1016/j.concog.2015.05.014

Rivers, M. L., Fitzsimmons, C. J., Fisk, S. R., Dunlosky, J., & Thompson, C. A. (2021). Gender differences in confidence during number-line estimation. *Metacognition and Learning*, *16*, 157-178. https://doi.org/10.1007/s11409-020-09243-7

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, *1*, e17. https://doi.org/10.1017/pen.2018.16

Sacher, M., Landré, L., & Taconnat, L. (2015). Age-related differences in episodic feeling-of-knowing arise from differences in memory performance. *Memory*, *23*(2), 119–126. https://doi.org/10.1080/09658211.2013.870210

Sassenberg, U., & Van Der Meer, E. (2010). Do We Really Gesture More When It Is More Difficult? *Cognitive Science*, *34*(4), 643–664. https://doi.org/10.1111/j.1551-6709.2010.01101.x

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of personality and social psychology*, *60*(3), 362.

Schachter, S., Rauscher, F., Christenfeld, N., & Crone, K. T. (1994). The Vocabularies of Academia. *Psychological Science*, *5*(1), 37–41. https://doi.org/10.1111/j.1467-9280.1994.tb00611.x

Schnadt, M. J., & Corley, M. (2006). *The Influence of Lexical, Conceptual and Planning Based Factors on Disfluency Production*.

Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture*. Ponsen & Looijen.

Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, *35*, 53–65. https://doi.org/10.1016/j.concog.2015.04.015

Smith, V. L., & Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, *32*(1), 25–38. https://doi.org/10.1006/jmla.1993.1002

Song, H., & Schwarz, N. (2008). Fluency and the Detection of Misleading Questions: Low Processing Fluency Attenuates the Moses Illusion. *Social Cognition*, *26*(6), 791–799. https://doi.org/10.1521/soco.2008.26.6.791

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94. https://doi.org/10.1016/j.jml.2005.02.003

ter Bekke, M., Drijvers, L., & Holler, J. (2024a). Gestures speed up responses to questions. *Language, Cognition and Neuroscience*, *39*(4), 423–430. https://doi.org/10.1080/23273798.2024.2314021

ter Bekke, M., Drijvers, L., & Holler, J. (2024b). Hand Gestures Have Predictive Potential During Conversation: An Investigation of the Timing of Gestures in Relation to Speech. *Cognitive Science*, *48*(1), e13407. https://doi.org/10.1111/cogs.13407

Tree, J. E. F. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, *34*(6), 709–738. https://doi.org/10.1006/jmla.1995.1032

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130292. https://doi.org/10.1098/rstb.2013.0292

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package ‘ggplot2’. *Create elegant data visualisations using the grammar of graphics. Version*, *2*(1), 1-189. https://github.com/tidyverse/ggplot2

Wokke, M. E., Achoui, D., & Cleeremans, A. (2020). Action information contributes to metacognitive decision-making. *Scientific reports*, *10*(1), 3632. https://doi.org/10.1038/s41598-020-60382-y

Yoon, S. O., & Brown-Schmidt, S. (2024). Partner-Specific Adaptation in Disfluency Processing. *Cognitive science*, *48*(8), e13490. https://doi.org/10.1111/cogs.13490

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*(2), 229–241. https://doi.org/10.3758/s13421-012-0255-8