Kısa Y.D., Goldin-Meadow, S. & Casasanto, D. (2024). Gesturing during disfluent speech: A pragmatic account. Cognition, 250, 105855. https://doi.org/10.1016/j.cognition.2024.105855.

Gesturing during disfluent speech: A pragmatic account

Yağmur Deniz Kısa[a], Susan Goldin-Meadow[b], and Daniel Casasanto[c]

[a]Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[b]Department of Psychology, University of Chicago, Chicago IL, USA

[c]Department of Psychology, Cornell University, Ithaca NY, USA

Author Note: This research was supported in part by a James S. McDonnell Foundation Scholar Award (#220020236) and a National Science Foundation grant (BCS #125710) to DC. Thanks to Kyle Jasmin, Defu Yap, and Cordelia Achen for help with data coding; to Laura Staum Casasanto and the members of the Goldin-Meadow lab and the Experience and Cognition Lab for discussion; to Barbara Tversky for making her lab rooms available. This study was included in Yağmur Deniz Kısa's PhD thesis at the University of Chicago. A preliminary report on these results was presented at the 2021 Association for Psychological Science Virtual Convention, at the 9th Conference of the International Society for Gesture Studies and at the 44th Annual Conference of the Cognitive Science Society. Data are available on the Open Science Framework (https://osf.io/yp3tr/). Correspondence should be addressed to Daniel Casasanto (casasanto@alum.mit.edu), Cornell University, Martha Van Rensselaer Hall, Ithaca, NY 14853.

Kısa Y.D., Goldin-Meadow, S. & Casasanto, D. (2024). Gesturing during disfluent speech: A pragmatic account. Cognition, 250, 105855. https://doi.org/10.1016/j.cognition.2024.105855.

Abstract

People are more likely to gesture when their speech is disfluent. Why? According to an influential proposal, speakers gesture when they are disfluent because gesturing helps them to produce speech. Here, we test an alternative proposal: People may gesture when their speech is disfluent because gestures serve as a pragmatic signal, telling the listener that the speaker is having problems with speaking. To distinguish between these proposals, we tested the relationship between gestures and speech disfluencies when listeners could see speakers' gestures and when they were prevented from seeing their gestures. If gesturing helps speakers to produce words, then the relationship between gesture and disfluency should persist regardless of whether gestures can be seen. Alternatively, if gestures during disfluent speech are pragmatically motivated, then the tendency to gesture more when speech is disfluent should disappear when the speaker's gestures are invisible to the listener. Results showed that speakers were more likely to gesture when their speech was disfluent, but only when the listener could see their gestures and not when the listener was prevented from seeing them, supporting a pragmatic account of the relationship between gestures and disfluencies. People tend to gesture more when speaking is difficult, not because gesturing facilitates speech production, but rather because gestures comment on the speaker's difficulty presenting an utterance to the listener.

Keywords: Gesture, Speech disfluency, Pragmatics, Lexical Retrieval Hypothesis, Pragmatic Signaling Hypothesis

Wordcount: 7731

Introduction

Why do people gesture when they speak? *When* speakers gesture can offer clues as to *why* they gesture. People are more likely to gesture when their speech is disfluent, compared to when their speech is fluent (Akhavan et al., 2016; Butterworth & Beattie, 1978; Ragsdale & Silvia, 1982). Why do gestures accompany disfluencies in speech? And what does this tendency tell us about the mechanisms by which gestures arise and the functions that they serve?

It is clear that gestures are, in part, produced for the listener: Whether speakers' gestures are visible to their listener affects how they gesture (see Bavelas & Healing, 2013, for a review). But the fact that speakers are more likely to gesture when their speaking is disfluent has been interpreted as evidence that gestures during disfluent speech must serve some function for the speaker, rather than for the listener.

According to a long-standing idea, which we will call the *Speech Facilitation Hypothesis*, people gesture when they speak because gestures help them produce speech (see for example Butterworth & Hadar, 1989; Hadar, 1989; Rauscher et al., 1996; Ravizza, 2003). The Speech Facilitation Hypothesis, as we call it here, includes different versions. One influential version is the Lexical Retrieval Hypothesis, which proposes that speech is facilitated *only* by gestures that carry semantic content, such as iconics and metaphorics (i.e., semantic gestures; e.g., Hadar, 1989; Rauscher et al., 1996). Other versions of the Speech Facilitation Hypothesis, by contrast, suggest that speech is facilitated by gestures that do not carry semantic content, such as beats (i.e., non-semantic gestures; Ravizza, 2003; Lucero et al., 2014). We test both versions of the Speech Facilitation Hypothesis here.

According to any version of the Speech Facilitation Hypothesis, people should gesture more when they are disfluent because those gestures help to resolve speech difficulties, thus

facilitating speech production. The Speech Facilitation Hypothesis predicts that when speakers are prevented from gesturing, they should experience more speech difficulties. However, contrary to this prediction, Kısa, Goldin-Meadow and Casasanto (2022) reviewed five decades of research on the relationship between gesture prevention and speech production and concluded that there is no reliable evidence that preventing gestures impairs speaking. This conclusion calls into question the primary source of empirical support for the Speech Facilitation Hypothesis, and challenges the proposal that gestures help resolve speech difficulties.

If gestures do not help resolve speech difficulties, then why are speakers more likely to gesture when they are disfluent? Here, we test an alternative proposal: Speakers gesture when their speech is disfluent because gestures serve as a pragmatic signal to the listener, commenting on the speaker's difficulties with presenting an utterance. We call this the *Pragmatic Signaling Hypothesis*.

According to Herbert Clark's (1996) theory of communication, speakers signal through two tracks simultaneously. In the *primary track*, they refer to the "official business" – the topic being talked about. In the *collateral track*, speakers comment on their performance of speaking. One situation when speakers should comment on the act of speaking is when they encounter problems with speech production – that is, delays or mistakes in presenting the official business. When speakers need extra time to plan their utterance or when they say the wrong word or phrase, they should give an account of the problem. There are many reasons why speakers might acknowledge that they are deviating from their expected performance of the utterance. Commenting on speech problems can give the listener information about the speaker's production plan and ensure successful coordination in conversation timing (Holler & Levinson, 2019). Commenting on speech problems is also motivated by conversational partners' social

expectations: Conversation is a joint activity and speakers would be violating the principle of being cooperative if they do not acknowledge their deviation from the role they commit to as a speaker — presenting an utterance (Clark, 1996). Accordingly, speakers use filled pauses (e.g., "um" and "uhh"; Clark & Fox Tree, 2002) as pragmatic signals to account for interruptions in their speech.

Like filled pauses, gestures could also serve as pragmatic signals, commenting on the speaker's difficulty with presenting an utterance (Clark, 1996). People may be more likely to gesture when they are disfluent, compared to when they are fluent, because gestures can comment on problems with presenting an utterance. This can happen in many ways. For example, gestures can signal the intent to continue speaking during an interruption, allowing the speaker to 'hold the floor' during a disfluency (Butterworth & Hadar, 1989; Duncan, 1972). Beyond floor holding, however, gestures can also foreshadow an upcoming interruption, acknowledge an ongoing interruption, or signal the speaker's commitment to a fluent re-start – these are all ways in which gestures can signal deviations from the speaker's "official business," and assure the listener that they intend to fulfil their communicative expectations.

The present study evaluated the Speech Facilitation Hypothesis and the Pragmatic Signaling Hypothesis as accounts of why speakers gesture more when their speech is disfluent. To distinguish between these hypotheses, we tested the relationship between gestures and speech disfluencies (i.e., within-phrase pauses, repeats or repairs of words or phrases, and filled pauses) when listeners could see speakers' gestures and when they were prevented from seeing their gestures. If gesturing during disfluent speech is pragmatically motivated, then when the listener cannot see the speaker, the speaker's motivation to gesture during disfluent speech should weaken or disappear. That is, according to the Pragmatic Signaling Hypothesis, people should be

more likely to gesture when they are disfluent than when they are fluent, only (or primarily) when the listener can see their gestures and potentially receive the pragmatic signal. By contrast, if gesturing during disfluent speech is motivated by facilitating speech production, then visibility should not matter because gestures should help speakers to resolve speech difficulties whether or not the listener can see them. That is, according to the Speech Facilitation Hypothesis, people should be more likely to gesture when they are disfluent whether or not the listener can see their gestures.

To preview our findings, speakers were more likely to gesture when their speech was disfluent, but *only* when the listener could see their gestures and *not* when the listener was prevented from seeing them, supporting the Pragmatic Signaling Hypothesis. These results fail to provide evidence for any version of the Speech Facilitation Hypothesis since gestures were more likely to occur during disfluent speech only when the listener could see the speaker's gestures both for semantic gestures and for non-semantic gestures. Gesturing during disfluent speech seems to be pragmatically motivated rather than being motivated by speakers' needs to help their own speech production.

Method

Transparency and openness

We used a pre-existing corpus of speech and gesture that was collected between 2005-2008 at Stanford University, for a study approved by Stanford University's Institutional Review Board. The corpus was designed to elicit gestures when participants told stories with literal and metaphorical spatial content, and when their gestures were visible and not visible. The sample size and the experimental manipulations (e.g., how the stories were constructed in terms of

semantic content, how gesture visibility was manipulated) for the present study were determined by the corpus we used.

Analyses of the clauses and gestures were reported in Yap and colleagues (2018). Analyses of the disfluencies in the gestures-visible condition were reported in Kısa and colleagues (2022). No analysis of disfluencies in the gestures-not-visible condition have been reported previously, nor have analyses of the relation between gestures and disfluencies in any of the visibility conditions.

The processed data needed for the analyses reported here, the R code to reproduce the analyses and the figures, the manuals used for gesture and disfluency coding, and example story transcripts and videos, are all available on the Open Science Framework (https://osf.io/yp3tr/). The study design, hypotheses, and analysis plan were not preregistered publicly. However, the study design, hypotheses, and analysis plan were proposed at a meeting of the Experience and Cognition Lab, Cornell University. The project proposal presented at this meeting can be found on the Open Science Framework (https://osf.io/rcvu2/).

We report how we determined all data exclusions and all measures in the study. The measures we used for gesture and disfluency coding were based on previous studies: The gesture coding was based on the gesture coding from Yap and colleagues (2018); the disfluency coding was based on the disfluency coding from Kısa and colleagues (2022).

Source corpus

The corpus we analyzed came from 56 Stanford University undergraduates, recruited in pairs, who participated for course credit after giving informed consent. Participants were told that the experiment was about storytelling. They took turns studying written stories, each for 60 seconds, and then retelling the stories to their partners. Participants were told to retell the stories

as accurately as possible because their partner would be quizzed on the content of the stories. All stories were written in the second person (e.g., "You're testing some new model rockets"), but participants were asked to retell the stories in the first person (e.g., "I'm testing some new model rockets") as if retelling their own experiences. There were 15 brief stories in total, each 50-100 words (see https://osf.io/kt28g/ for example story transcripts for each of the 15 stories and see https://osf.io/yp3tr/ for example videos of story retellings). After starting with a warm-up story, each participant re-told 6 stories in randomized order. Each story-telling session lasted 20-30 minutes.

Each pair of participants was assigned to one of two visibility conditions: gestures visible or gestures not visible. In the gestures-visible condition, participants were seated facing one another across a table. In the gestures-not-visible condition, the listener was blindfolded and the participants were separated by an opaque barrier on the table's surface occluding gesture space.

Coding

Clause coding

We used Yap and colleagues (2018)'s coding of the clauses in the source corpus. Participants' audio recordings of the stories were transcribed verbatim. Participants retold a total of 336 stories. The video recording for 3 of the stories are missing so, for further analyses, we worked with 333 stories in total. Yap and colleagues (2018) parsed transcriptions of participants' audio recordings into clauses. Participants produced a total of 3534 spoken clauses (Gestures-visible: 1936; Gestures-not-visible: 1598). The number of words people produced per clause was very similar across the visibility conditions: Both the participants in the gestures visible condition and those in the gestures not visible condition produced an average of 9 words per

clause (gestures visible: M=8.60, SD=4.32; gestures not visible: M=9.04, SD=4.16), suggesting that speech production was similar across the visibility conditions.

A total of 360 spoken clauses were excluded from further analyses: (i) Gestures were not codable for 108 of these clauses because the hands were occluded from view and (ii) disfluencies were not codable for 252 of these clauses due to low audio quality of the recording. As a result, a total of 3174 clauses were included in further analyses.

Speech disfluency coding

Coder 1 recorded whether speech disfluencies were present for each clause, using only the audio with no video. Doing the coding with no video ensured that speech disfluencies were coded without any knowledge of the gestures. A clause was categorized as containing speech disfluencies if it included unfilled pauses, repeats, repairs, or filled pauses (see the disfluency coding manual on https://osf.io/6bjsg/ for more details). We did not have any specific hypotheses about different disfluency types; therefore a clause was considered to be disfluent if it included *any* of the four disfluency types. Unfilled pauses included silences that could be associated with word retrieval difficulty: silences within words, silences between simple modifiers and heads, between heads and simple complements, between compound verbs and compound noun phrases (e.g., “fall [pause] down”). Silences were not considered a pause if they occurred after a discourse marker (e.g., “well”, “you know”, “but”, etc.), after reporting verbs (e.g., “I think”), or between phrases. Repeats included repetition of words that were exactly the same as what came before (e.g., “I went to the [pause] to the store”). Repairs included modifications in speech where what came after was meant to overwrite what came before (e.g., “I went to the shore [pause] to the store”). Filled pauses included “umm”s and “uhh”s.

Participants produced a total of 1905 disfluencies (Gestures visible: 988; Gestures not visible: 917). Speech disfluencies included 457 unfilled pauses, 232 repeats, 391 repairs, and 824 filled pauses. The number of disfluencies produced across the visibility conditions were similar: Participants in the gestures visible condition produced a total of 988 disfluencies (210 repairs, 124 repeats, 434 filled pauses and 210 unfilled pauses) and participants in the gestures not visible condition produced a total of 917 disfluencies (172 repairs, 108 repeats, 390 filled pauses and 247 unfilled pauses).

A clause was classified as Disfluency Present if it included at least one disfluency. Forty-two percent of the clauses (1327 out of 3174) had at least one disfluency associated with them. Coder 2 coded speech disfluencies for a randomly selected 10% of all stories and the intercoder agreement for whether a clause contained disfluencies was 93% (*Cohen's kappa* = .86, $z = 15.3$, $p < .001$).

Gesture coding

We used Yap and colleagues (2018)'s coding of the gestures in the source corpus. Yap and colleagues (2018) classified gestures into iconics, metaphorics, deictics, or emblems, according to McNeill's (1992) gesture categories (see the gesture coding manual used by Yap and colleagues, 2018, on https://osf.io/74bre/ for more details). Iconics are gestures that depict concrete things and/or actions,, such as making a holding shape to depict holding a cup; metaphorics are gestures that depict abstract things as if they are concrete things and/or actions (e.g. metaphorically holding ideas in hands when one says "on the one hand"); deictics are pointing gestures that involve the extension of a finger, hand or arm to indicate an entity; and emblems are gestures with a conventional meaning such as thumbs up (McNeill, 1992; Cartmill and Goldin-Meadow, 2016). We excluded deictics and emblems from our analyses (and only

focused on iconics, metaphorics, and beats) because deictics and emblems are known to serve clear communicative functions; the hypotheses we tested in the current study concern gestures that have been hypothesized previously to serve speaker-internal cognitive functions (iconics, metaphorics, and beats; see for example Rauscher et al., 1996; Ravizza, 2003; Lucero et al. 2014).

Yap and colleagues (2018)'s coding proceeded in two stages. In the first stage, using only the video with no audio, the stroke phase of each gesture was determined to be a beat or non-beat according to McNeill's (1992) beat filter. Gestures were classified as beats if they had (i) two movement phases, (ii) a relaxed handshape, and (iii) movement only within a single region of gesture-space. Gestures that included any features of other gesture types (e.g., iconics) were not classified as beats. Doing the initial gesture identification with no audio ensured that gestures were coded without any knowledge of the speech disfluencies.

In the second stage, using both audio and video, non-beat gestures were classified into iconics, metaphorics, deictics, or emblems, according to the gestures' forms and the accompanying speech, following McNeill's (1992) gesture categories. Participants produced a total of 3192 gestures (Gestures visible: 2012; Gestures not visible: 1180). Gestures in the source corpus included 446 iconics, 65 metaphorics, 2392 beats, 288 deictics and 1 emblem. As mentioned earlier, we focused on iconics, metaphorics, and beats.

A clause was classified as Gesture Present if it included at least one gesture (iconic, metaphoric, or beat). Forty-six percent of the clauses (1473 out of 3174) had at least one gesture associated with them.

Semantic and non-semantic gestures

We categorized each gesture coded by Yap and colleagues (2018) as either a *semantic gesture*, which conveyed semantic meaning and thus contributed to the official business of an utterance, or *a non-semantic gesture,* which did not convey semantic meaning. Our categorization did not involve any new coding of gesture types, but simply involved placing the already coded traditional gesture types into our novel semantic and non-semantic gesture categories. Consistent with standard practice, we categorized iconics and metaphorics as semantic gestures. Beats have traditionally been categorized as devoid of semantic meaning (McNeill, 1992). However, when analyzing the same corpus we used here, Yap and colleagues (2018) showed that many beat gestures reflect the spatial semantics of the utterances they accompany. For example, people tended to produce upward beat gestures when their speech implied upward motion (e.g., "my rocket went higher") and downward beat gestures when their speech implied downward motion (e.g., "the scuba diver went down"), more frequently than expected by chance. Following Yap and colleagues (2018), here we distinguish between two types of beat gestures: semantic beats that reflect spatial semantics, and non-semantic beats that do not reflect the spatial semantics of the accompanying speech.

The stories in the corpus were designed to elicit gestures during speech with spatial content (literal or metaphorical), and each story implied motion or extension in one of four spatial directions: upward, downward, right, or left. Within each story, some clauses expressed spatial direction (directional clauses, e.g., "my rocket went higher"), whereas other clauses did not express any spatial direction (non-directional clauses, e.g., "I'm testing some new model rockets").

Yap and colleagues (2018) coded the direction of the stroke for each gesture using silent videos as upward, downward, leftward, rightward, or other. Following Yap et al.'s (2018)

coding, here we classified beat gestures as *semantic beats* if the direction of the beat gesture (e.g., upward) was the same as the direction implied by the accompanying clause (for directional clauses) or the same as the overall direction implied by the story (for non-directional clauses). We classified beat gestures as *non-semantic beats* if the direction of the beat gesture was different from the direction implied by the accompanying clause (for directional clauses) or from the overall direction implied by the story (for non-directional clauses).

Having classified the beats as described above, our category of semantic gestures included iconics, metaphorics, and semantic beats, and our category of non-semantic gestures included only non-semantic beats. Overall, participants produced a total of 1356 semantic gestures, including 446 iconics, 65 metaphorics, and 845 semantic beats, and a total of 1521 non-semantic gestures (i.e., non-semantic beats). Twenty-six of the beat gestures were excluded because either they could not be categorized as semantic or non-semantic since multiple spatial directions were implied by the accompanying clause, or because the coder was unsure about coding the gesture as a beat.

A clause was classified as Non-semantic Gesture Present if it included non-semantic gestures only (i.e., clauses with non-semantic beats, and no other gestures during the clause; a total of 608 clauses). A clause was classified as Semantic Gesture Present if it included semantic gestures only (i.e., clauses with iconics, metaphorics, or semantic beats, and no other gestures during the clause; a total of 559 clauses). A clause was classified as No Gesture Present if it did not include semantic or non-semantic gestures (a total of 1710 clauses). A clause was classified as Mixed Type if it contained both semantic and non-semantic gestures (a total of 297 clauses). These Mixed Type clauses were included in the overall analyses of the relationship between disfluency and gesture production; they were excluded from the separate analyses of this

relationship for semantic gestures vs. non-semantic gestures in order to ensure that these categories of gestures were independent of each other.

Analyses

We conducted all analyses by fitting generalized linear mixed-effect models, using R (R Core Team, 2020; see the R code for the analyses on https://osf.io/t9yjn/), the glmer() function in the lme4 library (Bates et al., 2015). We used a "maximal" random effect structure justified by our design (Barr et al., 2013). We treated Subject (N = 53) and Story (N = 15) as random effects, including random intercepts for both in analyses, since our outcome variables (e.g., Gesture Presence) are likely to vary across different subjects and stories. Subjects are likely to vary idiosyncratically also in their sensitivity to our fixed factor, Disfluency Presence. Therefore, we also included random slopes for our within-subject fixed factor (i.e., Disfluency Presence), allowing subjects' likelihood to gesture to vary differentially based on our fixed factor (e.g., subjects could be affected differently by Disfluency Presence in their production of gestures). We chose to model both our outcome variables (i.e. Gesture Presence, Semantic Gesture Presence, Non-Semantic Gesture Presence) and Disfluency Presence as binary variables (present or absent), rather than counts (number of gestures or disfluencies), since only a small proportion of our clauses contained more than one gesture (only 10 percent contained more than one semantic gesture and only 12 percent contained more than one non-semantic gesture) or more than one disfluency (13 percent of all clauses). We used likelihood ratio tests (LRTs) to test for the omnibus interaction effects. We estimated group means and performed our planned contrasts using the emmeans() function in the emmeans package (Lenth et al., 2018) in R. We reported Odds Ratios (OR) in the response scale (probability) as a measure of effect size for the simple effects.

Results

Effect of disfluency and visibility on gesture production

To compare Gesture Presence across experimental conditions for each clause (N = 3174), we used mixed effects logistic regressions with Gesture Present and No Gesture Present as the binary outcomes for a clause[1].

We first tested whether people were more likely to gesture when they were disfluent, compared to when they were fluent, when their gestures were visible. Participants whose gestures were visible were more likely to gesture during disfluent clauses than fluent clauses (z= 4.56, p< .001, *OR*=2.24, 95%CI [1.58, 3.17]; see the fluent and disfluent columns on the left in Figure 1). This result replicates previous studies (Akhavan et al., 2016; Butterworth & Beattie, 1978; Ragsdale & Silvia, 1982), showing a positive relationship between gesture production and speech disfluencies when gestures are visible, an outcome predicted by both the Speech Facilitation Hypothesis and the Pragmatic Signaling Hypothesis.

Next, we tested whether people gestured more during disfluent clauses when their gestures were *not* visible. For participants whose gestures were not visible, there was no significant effect of disfluency on gesture production (z = 1.21, p= .22, *OR*=1.25, 95%CI [0.87, 1.80], see the fluent and disfluent columns on the right in Figure 1). These results are not consistent with the Speech Facilitation Hypothesis, according to which people should be more likely to gesture when they are disfluent whether or not the listener can see their gestures. Rather,

[1] R syntax for the omnibus interaction model: gesture presence ~ disfluency presence * gesture visibility + (1 + disfluency presence | subject) + (1 | story)

these results provide evidence for the Pragmatic Signaling Hypothesis: The absence of a positive relationship between gesture production and disfluent speech when these gestures cannot be seen supports the proposal that gestures serve as pragmatic signals during disfluent speech, meant to be seen by a listener.

The Pragmatic Signaling Hypothesis additionally predicts that visibility (Gesture Visible, Gesture Not Visible) and disfluency (Disfluent Clause, Fluent Clause) will interact to predict gesture production (Gesture Present, No Gesture Present). If gestures during disfluent speech are pragmatically motivated, people should be more likely to gesture during disfluent speech *only* when their gestures are visible – resulting in a significant difference between the effect of disfluency on gesture production in the gestures visible condition and the *absence* of an effect of disfluency on gesture production in the gestures not visible condition. Results showed that visibility and disfluency interacted to predict gesture production: People were more likely to gesture during disfluent clauses than during fluent clauses *only* when their gestures were visible ($\chi 2(1) = 5.20$, $p = .023$) – providing support for the Pragmatic Signaling Hypothesis.

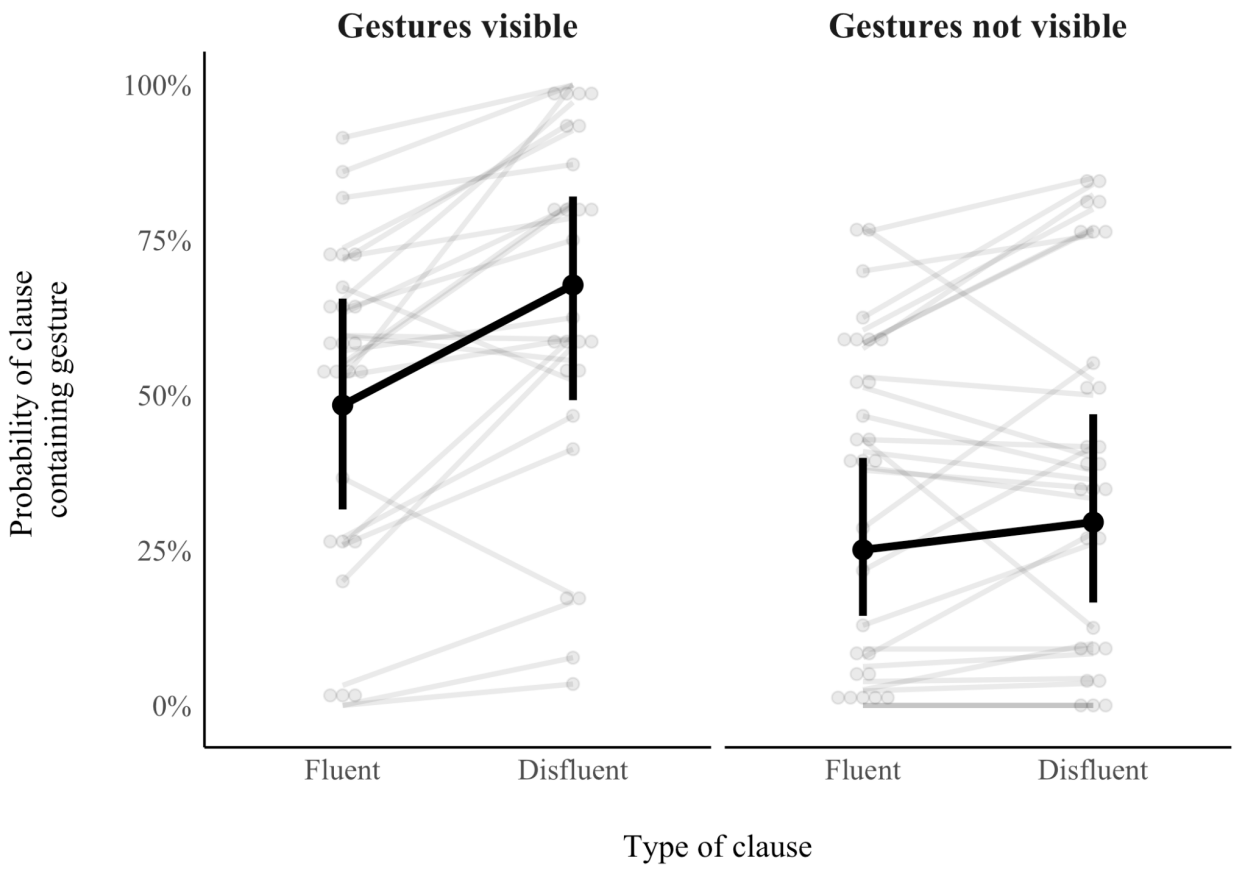

Figure 1. Results showing the probability of clauses containing gestures during fluent and disfluent speech (# of clauses that contain gestures / # of all clauses), for participants whose gestures were visible (left) and participants whose gestures were not visible (right) to their interlocutor. Black points represent the estimated group means and black error bars around these points represent the estimated 95% confidence intervals (asymp.LCL and asymp.UCL) for the mixed model, calculated using the emmeans() function in the emmeans package in R. Gray lines represent the data summary with the means for individual participants.

Testing the Speech Facilitation Hypothesis in semantic vs. non-semantic gestures

Our main analysis collapsed over both semantic and non-semantic gestures. Yet, according to some versions of the Speech Facilitation Hypothesis, speech is facilitated only by gestures that carry semantic content, like words do (e.g., Hadar, 1989; Rauscher et al., 1996). Other versions of the Speech Facilitation Hypothesis, by contrast, suggest that speech is facilitated by beat gestures (Lucero et al., 2014) or by motor actions with no semantic content (Ravizza, 2003). Is it

possible that support for the Speech Facilitation Hypothesis could be found only for semantic gestures (as suggested by Rauscher et al., 1996), or only for non-semantic gestures (as suggested by Lucero et al., 2014)? To explore this possibility, we tested for an effect of disfluency on gesture production in non-semantic and semantic gestures, separately.

Non-semantic gestures

Nineteen percent of the clauses (608 out of 3174) had only non-semantic gestures associated with them. To compare Non-semantic Gesture Presence across experimental conditions for each clause (N = 2318; 608 clauses with non-semantic gestures only, and 1710 with no gestures), we used mixed effects logistic regressions with Non-semantic Gesture Present and No Gesture Present as the binary outcomes for a clause[2]. We excluded clauses that contained semantic gestures from this analysis (856 clauses; 559 clauses with semantic gestures only, and 297 with both semantic and non-semantic gestures).

We first tested whether people were more likely to produce non-semantic gestures than no gestures when they were disfluent, compared to when they were fluent, when their gestures were visible. Participants whose gestures were visible were more likely to produce non-semantic gestures than no gestures during disfluent clauses, compared to fluent clauses (z=4.36, p< .001, *OR*=2.27, 95%CI [1.57,3.29]; see the fluent and disfluent columns on the left in Figure 2).

[2] R syntax for the omnibus interaction model: non-semantic gesture presence ~ disfluency presence * gesture visibility + (1 + disfluency presence | subject) + (1 | story). Note that the data in the model did not include clauses that have semantic gestures – neither clauses that only have semantic gestures nor clauses with both semantic and non-semantic gestures. Also note that the non-semantic gesture analysis presented here is not independent from the overall gesture analysis presented earlier, since both analyses include data from clauses with no gesture.

We next tested whether the link between producing non-semantic gestures and disfluencies is found even when speakers' gestures are not visible. If producing non-semantic gestures during disfluent speech is motivated by speaker-internal needs, as predicted by the Speech Facilitation Hypothesis, then people should be more likely to produce non-semantic gestures during disfluent speech even when the speaker's gestures are *not* visible. However, our results failed to provide support for the Speech Facilitation Hypothesis: For participants whose gestures were not visible, there was no significant effect of disfluency on non-semantic gesture production (z =0.88, p= .377, *OR*=1.20, 95%CI [0.80,1.82], see the fluent and disfluent columns on the right in Figure 2). These results suggest that producing non-semantic gestures during disfluent speech may *not* be motivated by speaker-internal needs. Rather, these findings are consistent with the Pragmatic Signaling Hypothesis: The tendency to produce non-semantic gestures during disfluent speech *disappears* when these gestures cannot be seen, suggesting that non-semantic gestures during disfluent speech are produced as pragmatic signals for the listener.

Finally, our results showed that visibility and disfluency interacted to predict gesture production: People were more likely to gesture during disfluent clauses than during fluent clauses *only* when their gestures were visible ($\chi 2(1) = 5.20$, p= .023). These results provide clear support for the Pragmatic Signaling Hypothesis as an explanation for why people are more likely to produce non-semantic gestures during disfluent speech: Non-semantic gestures are produced to serve as pragmatic signals for the listener during disfluent speech.

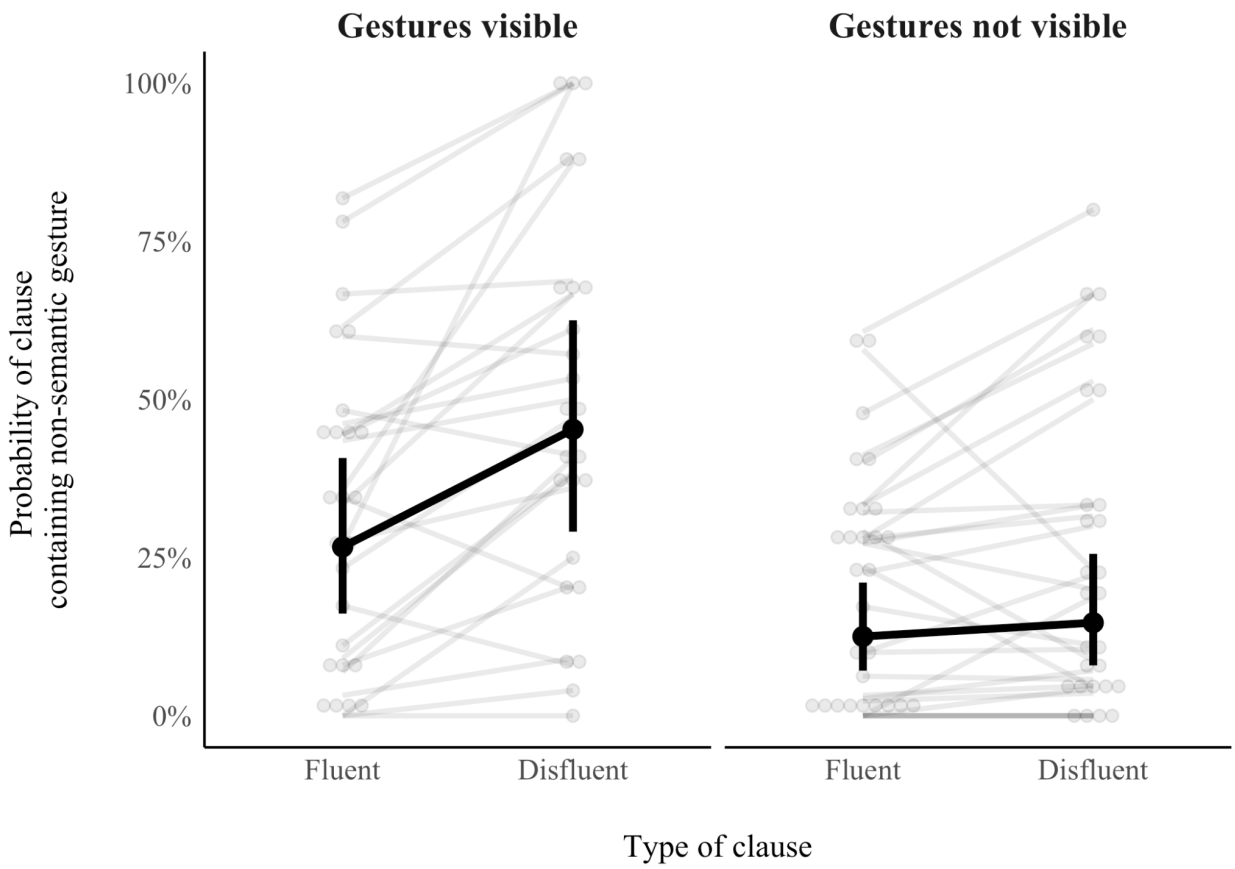

Figure 2. Results showing the probability of clauses containing non-semantic gestures (# of clauses that contain non-semantic gestures only / (# of clauses that contain non-semantic gestures only + # of clauses with no gestures) for fluent and disfluent clauses, for participants whose gestures were visible (left) and participants whose gestures were not visible to their interlocutor (right). Black points represent the estimated group means and black error bars around these points represent the estimated 95% confidence intervals (asymp.LCL and asymp.UCL) for the mixed model, calculated using the emmeans() function in the emmeans package in R. Gray lines represent the data summary with the means for individual participants.

Semantic gestures

Next we turned to semantic gestures – iconics, metaphorics and semantic beats. Is it possible that support for the Speech Facilitation Hypothesis could be found only for semantic gestures? To explore this possibility, we tested for an effect of disfluency on gesture production when gestures are visible and when gestures are not visible, separately. If producing semantic gestures during disfluent speech is motivated by speaker-internal needs, then people should be

more likely to produce semantic gestures regardless of whether their gestures are visible to a listener.

Eighteen percent of the clauses (559 out of 3174) had only semantic gestures associated with them – 189 clauses with iconic gestures only, 35 clauses with metaphoric gestures only, 286 gestures with congruent beats only, and 49 clauses with some mixture of these semantic gesture types. To compare Semantic Gesture Presence across experimental conditions for each clause (N = 2269; 559 clauses with semantic gestures only, and 1710 clauses with no gestures), we used mixed effects logistic regressions with Semantic Gesture Present and No Gesture as the binary outcomes for a clause[3]. We excluded clauses that contained non-semantic gestures for this analysis (905 clauses, total, containing 608 clauses with non-semantic gestures only; 297 clauses with both non-semantic and semantic gestures).

Participants whose gestures were visible were more likely to produce semantic gestures than to produce no gestures during disfluent clauses, compared to fluent clauses (z=2.22, p=.027, OR=1.63, 95%CI [1.06,2.52]; see the first two columns in Figure 3). We next tested whether the link between producing semantic gestures and disfluencies is found even when speakers' gestures are not visible. As was the case for non-sematic gestures, our results failed to provide support for the Speech Facilitation Hypothesis for semantic gestures: For participants whose

[3] R syntax for the omnibus interaction model: semantic gesture presence ~ disfluency presence * gesture visibility + (1 + disfluency presence | subject) + (1 | story). Note that the data in the model did not include clauses that have non-semantic gestures – neither clauses that only have non-semantic gestures nor clauses with both non-semantic and semantic gestures. Also note that the semantic gesture analysis presented here is not independent from the overall gesture analysis or the non-semantic gesture analysis presented earlier, since all of these analyses include data from clauses with no gesture.

gestures were *not* visible, there was no statistically significant evidence that people were more likely to produce semantic gestures than to produce no gestures during disfluent clauses, compared to fluent clauses (z=0.35, p=.722, OR=1.09, 95%CI [0.66,1.80]; see the last two columns in Figure 3). These results suggest that *neither* non-semantic gestures *nor* semantic gestures that speakers produce during disfluent speech may be motivated by speaker-internal needs. Rather, these findings are consistent with the Pragmatic Signaling Hypothesis, suggesting that semantic gestures during disfluent speech may be produced as pragmatic signals for the listener.

Finally, even though the link between disfluencies and semantic gesture production was found for participants whose gestures were visible, and not found for participants whose gestures were not visible (as reported above), the two-way interaction between visibility and disfluency was not a statistically significant predictor of semantic gesture production ($\chi2(1) = 1.44$, p= .230). The interaction between visibility and disfluency was not statistically significant for semantic gestures, but we found the same qualitative pattern for semantic gestures as for non-semantic gestures, consistent with the Pragmatic Signaling Hypothesis.

In summary, the separate analyses of non-semantic and semantic gestures failed to provide support for the Speech Facilitation Hypothesis: The relationship between speech disfluency and gesture production disappeared when gestures could not be seen, for both non-semantic and semantic gestures. The analysis of non-semantic gestures provides strong evidence for the Pragmatic Signaling Hypothesis, showing the same pattern of statistically significant results as we found in the analysis of all gesture types, combined. Although the analysis of semantic gestures showed the same qualitative pattern, the results provide only suggestive evidence that semantic gestures also serve this pragmatic function.

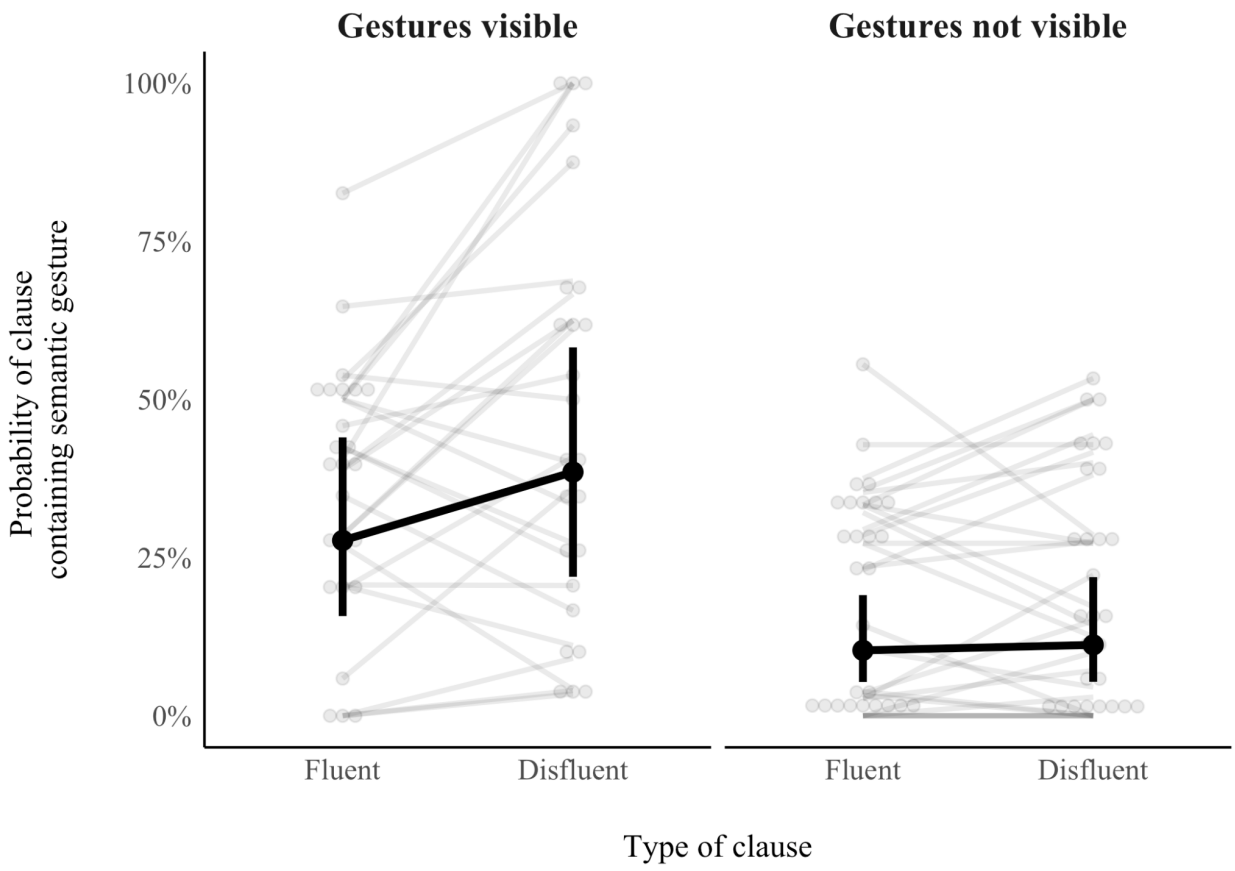

Figure 3. Results showing the probability of clauses containing semantic gestures (# of clauses that contain semantic gestures only / (# of clauses that contain semantic gestures only + # of clauses with no gestures) for fluent and disfluent clauses, for participants whose gestures were visible (left) and participants whose gestures were not visible to their interlocutor (right). Black points represent the estimated group means and black error bars around these points represent the estimated 95% confidence intervals (asymp.LCL and asymp.UCL) for the mixed model, calculated using the emmeans() function in the emmeans package in R. Gray lines represent the data summary with the means for individual participants.

Discussion

Why are people more likely to gesture when their speech is disfluent, compared to when their speech is fluent (Akhavan et al., 2016; Butterworth & Beattie, 1978; Ragsdale & Silvia, 1982)? Here we show that speakers gesture more when their speech is disfluent, but *only* when those gestures are visible to the listener. These results support the Pragmatic Signaling Hypothesis: Gestures produced during disfluent speech are pragmatic signals to the listener,

commenting on the speaker's difficulty presenting an utterance. These results fail to provide evidence for the Speech Facilitation Hypothesis: Gesturing during disfluent speech may not be motivated by facilitating speech production, given that the relationship between gesture and disfluency disappeared when gestures could not be seen by a listener. In other words, people tend to gesture more when their speech is disfluent, not because gesturing facilitates speech production, but rather because gestures serve as a pragmatic signal to the listener.

Are gestures during disfluent speech produced to facilitate speech?

Gestures that occur during difficult speech have been widely interpreted as evidence that many gestures are primarily produced to meet the speaker's needs (i.e., the Speech Facilitation Hypothesis; see for example Krauss & Hadar, 1999). One of the predictions of the Speech Facilitation Hypothesis is that when speakers are prevented from gesturing, they should experience more speech difficulties. However, in conflict with this prediction, Kısa and colleagues (2022) showed that there is no reliable evidence that preventing gestures impairs speaking, challenging the main source of empirical support for the Speech Facilitation Hypothesis.

The finding that gestures and speech disfluencies tend to co-occur has been interpreted as another source of empirical support for the Speech Facilitation Hypothesis (see for example Krauss & Hadar, 1999). In the present study, we challenged this interpretation, and tested an alternative explanation for the relationship between gestures and speech disfluencies. Our data are not compatible with the Speech Facilitation Hypothesis because gestures were not more likely to occur *whenever* speakers had speech disfluencies; rather, gestures were only more likely to occur when speakers had speech disfluencies *and* their gestures were visible. When the listener could not see the speaker, the speaker was not more likely to produce gestures during

disfluent clauses. If speakers gesture during disfluent speech to facilitate their own speech production, then they should gesture during speech problems whether or not their gestures are visible – contrary to the present findings.

Our main analysis focused on all gestures, but there are versions of the Speech Facilitation Hypothesis proposing that speech is facilitated only by gestures that carry semantic content (e.g., Hadar, 1989; Rauscher et al., 1996). Other versions of the Speech Facilitation Hypothesis, by contrast, suggest that speech is facilitated by gestures with no semantic content, such as beats (Lucero et al., 2014; Ravizza, 2003). In principle, it could be possible that we failed to find support for the Speech Facilitation Hypothesis in our main analysis because we collapsed all gesture types. To test this possibility, we analyzed semantic and non-semantic gestures separately to see whether people produce any of these gesture types due to speaker-internal needs. Separating the gesture types did not change the pattern of results: For both semantic and non-semantic gestures, gestures were more likely to occur only when speakers had speech disfluencies *and* their gestures were visible. When people's gestures were *not* visible to the listener, the relationship between disfluencies and gesture production was not found for semantic gestures *or* for non-semantic gestures. These results fail to provide evidence for either version of the Speech Facilitation Hypothesis. Gesturing during disfluent speech seems to be communicatively motivated for both semantic and non-semantic gestures, rather than being motivated by speakers' needs to help their own speech production.

The communicative motivations for gesturing during difficult speech are likely to be more widespread than the cases we discuss here. People are more likely to gesture not only when they are disfluent, but also when speaking is difficult but no disfluencies are produced. People are more likely to gesture when they are about to utter a low probability word; when they use a

less preferred syntactic construction; or when they are in a tip-of-the-tongue state (Beattie & Shovelton, 2000; Cook, Jaeger, & Tanenhaus, 2009; Holler, Turner, & Varcianna, 2013). It is possible that people gesture in these cases to help their own speech. Alternatively, just like gestures during disfluent speech, gestures produced during other speech difficulties may also be designed for the listener (see for example Holler, Turner, & Varcianna, 2013; Kısa, 2022).

Pragmatic signaling during disfluent speech: Beyond holding the floor

In line with the Pragmatic Signaling Hypothesis, many researchers argue that speakers use gestures to hold the floor during a pause, to prevent the listener from interrupting their turn (Butterworth & Hadar, 1989; Duncan, 1972) – this is one way in which speakers can use gestures as a pragmatic signal during disfluencies. Indeed, gesturing during pauses may sometimes be motivated by the need to hold the floor: People are more likely to gesture when they pause than when they utter a word *only* when they are in a conversation involving listener interruption, and not when they produce a monologue without any listener interruptions (Beattie & Aboudan, 1994; see Nobe, 2000, for a replication).

It is possible that some of the gestures during disfluent speech that we observed in the current study, specifically some gestures during pauses, may have been motivated by the need to hold the floor. However, holding the floor is not likely to be the only pragmatic motivation leading to gesturing during disfluencies in our study. First, the interactional context in the current study involved a smaller number of listener interruptions compared to a spontaneous dialogue: Pairs of participants took turns telling stories to each other and the listeners could interrupt the speaker/storyteller to ask questions on the story, since the listeners were quizzed on the stories afterwards. This set-up created a sequence of long turns by the storytellers, but these turns were at times interrupted by the listener asking questions, and listeners also often backchanneled. This

interactional setting is different from a dialogue where the speaker and the listener roles are constantly negotiated and there is a greater need to hold the floor. Second, unlike pauses, most of the disfluencies in our sample (repairs, repeats and fillers) rarely license turn interruptions, so they are not likely to require holding the floor. Furthermore, the pauses that we included occurred in the middle of a phrase, where listeners are unlikely to interrupt the speaker. Holding the floor is thus not likely to explain all (and perhaps not any) of the gestures during the disfluencies in our study. But all deviations from fluent speech in our study interrupted the official business of talking, and speakers may have used gestures to signal this deviation.

Why might speakers use gestures to signal deviations from the official business, even when they are not at risk of being interrupted? Deviations from an utterance plan are also deviations from the conversation timing plan – and commenting on speech problems can give the listener information about the speaker's production plan and thus ensure successful coordination in conversation timing (Holler & Levinson, 2019). Additionally, deviations from presenting an utterance are also deviations from the role a speaker commits to by being in a joint action of conversing. Speakers commit to presenting an utterance when they have the turn, and listeners expect them to continue to commit to this role and give an account of any deviation from it (Clark, 1996). For these reasons, gesturing during disfluent speech might not only have the specific motivation to ensure not losing your turn when your speech stops at a potential turn boundary, it might also have a more general motivation to comment on problems with speaking whenever your speech is drifting away from the official business.

Relying on previous work on how gestures are used during disfluent speech and on our own anecdotal observations of the current data, we speculate about various ways in which speakers can use gestures to comment on problems with speaking, beyond holding the floor:

Gestures can foreshadow an upcoming interruption; acknowledge an ongoing interruption; or signal speakers' commitment to a fluent re-start. Previous work showed that people tend to suspend their gestures (e.g., going into a hold) before they suspend their speech, suggesting that speakers could use gestures to signal an upcoming interruption with speaking (Seyfeddinipur & Kita, 2001). Other work shows that people tend to hold their gestures when producing a disfluency marker, suggesting that speakers could use gestures to acknowledge an ongoing interruption (Graziano & Gullberg. 2018). And other work shows that movements, including gestures, are more likely to occur during the first word following a speech disfluency (Dittmann & Llewellyn, 1969), suggesting that speakers could use gestures to signal their commitment to a fluent re-start once a disfluency is resolved. However, none of this previous work tested whether these gestures are designed as pragmatic signals for the listener. Future work could look at whether the previously studied temporal relationships between gestures and disfluencies appear *only* when gestures are visible to a listener, when they can serve as signals commenting on speaker's problems with speaking.

Which gestures serve as pragmatic signals during disfluent speech?

In principle, all types of gestures that are produced when people speak could serve as pragmatic signals commenting on problems with speaking. Accordingly, both the gestures that contribute to the official business of an utterance (semantic gestures) and the gestures that do not contribute to the official business (non-semantic gestures) could serve as pragmatic signals commenting on the process of speaking.

To determine which kinds of gestures served this pragmatic function, we tested which gestures were more likely to be produced during disfluent speech selectively when the gestures were visible. For non-semantic gestures, we found that disfluent clauses were more likely to have

a gesture than to have no gesture (compared to fluent clauses), but *only* when those gestures were visible. This pattern suggests that producing non-semantic gestures during disfluent speech is pragmatically motivated. For semantic gestures, we found a qualitatively similar pattern. When semantic gestures were visible, disfluent clauses were significantly more likely to have a gesture than to have no gesture, compared to fluent clauses. When semantic gestures were not visible, this relationship between gesture production and speech disfluency disappeared; however, the effect of visibility on the relationship between semantic gesture production and speech disfluency was not statistically significant. Overall, these results provide strong evidence that non-semantic gestures serve as pragmatic signals commenting on problems with speaking, and provide suggestive evidence that semantic gestures may also serve this pragmatic function.

Gesturing for the listener

People gesture, in part, for the listener. Speakers modify their gestures when their gestures are visible to an interlocutor, compared to when their gestures are not visible (see Bavelas & Healing, 2013, for a review). For example, people's gestures are bigger and less redundant with speech when their gestures are visible to an interlocutor (Bavelas et al., 2008) – showing that the form and semantic content of gestures are, at least in part, designed for the listener.

Gestures are also designed as pragmatic signals for the listener, commenting on the act of speaking (see Kendon, 2017, for a review). For example, speakers are more likely to use beat gestures when they depart briefly from the narrative when their gestures are visible, compared to when their gestures are not visible – suggesting that speakers design beat gestures as pragmatic signals to comment on discourse structure (McNeill, 1992; Alibali et al., 2001). Speakers are also more likely to use interactive gestures, which are gestures that comment on some aspect of

conversing with another person, when engaged in a dialogue with the interlocutor, compared to when engaged in a monologue where the interlocutor merely listens (Bavelas et al., 1995).

Here, we provided evidence for a novel hypothesis about speakers' use of gestures as pragmatic signals for the listener. Just like filled pauses in speech serve as pragmatic signals that account for interruptions in speech (Clark & Fox Tree, 2002), gestures can give the listener an account of the speaker's production plan by, for example, foreshadowing or acknowledging speech difficulties.

The listener's uptake of gesture's pragmatic meanings

Once we know more about the specific ways in which gestures comment on problems with speaking, we can then also ask whether listeners pick up on this commentary. Here, we showed that speakers produce gestures during disfluent speech primarily when the listener can see them, but the speaker's sensitivity to the listener does not guarantee that the listener will glean meaning from their gestures. Future work is needed to know whether listeners glean the pragmatic messages speakers intend to convey with the gestures that accompany disfluent speech.

Conclusions

People gesture when they experience difficulties speaking, and this well-established pattern has been interpreted as providing an answer to the question of why people gesture when they speak. Specifically, this pattern has been interpreted as evidence that gesturing helps speakers find the right words. Yet, the results of the present study are incompatible with this explanation of gesturing during disfluent speech. Rather than facilitating speech production, gesturing during disfluencies appears to serve as a pragmatic signal to the listener, commenting on the act of speaking.

References

Akhavan, N., Göksun, T., & Nozari, N. (2016). Disfluency production in speech and gesture. In Proceedings of the 38th annual meeting of the cognitive science society.

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255–278.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. Personality and social psychology bulletin, 21(4), 394–405.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. Discourse processes, 15(4), 469–489.

Bavelas, J. B., Jennifer, G., Chantelle, S., & Danielle, P. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. Journal of Memory and Language, 58, 495–520.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. Gesture, 13(1), 63-92.

Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental inves- tigation of the effects of changing social context on their precise temporal relationships. Semiotica, 99(3-4), 239–272.

Beattie, G., & Shovelton, H. (2000). Iconic hand gestures and the predictability of words in context in spontaneous speech. British Journal of Psychology, 91(4), 473–491.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In Recent advances in the psychology of language (pp. 347–360). Springer.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: a reply to mcneill. Psychological review, 96(1), 168–174.

Clark, H. H. (1996). Using language. Cambridge university press.

Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. Cognition, 84(1), 73-111.

Cook, S. W., Jaeger, T. F., & Tanenhaus, M. (2009). Producing less preferred structures: More gestures, less fluency. In Proceedings of the annual meeting of the cognitive science society (Vol. 31).

Dittmann, A. T., & Llewellyn, L. G. (1969). Body movement and speech rhythm in social conversation. Journal of personality and social psychology, 11(2), 98.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. Journal of personality and social psychology, 23(2), 283.

Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in psychology*, *9*, 879.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communi- cation. Trends in Cognitive Sciences, 23(8), 639–652.

Holler, J., Turner, K., & Varcianna, T. (2013). It's on the tip of my fingers: Co-speech gestures during lexical retrieval in different social contexts. Language and Cognitive Pro- cesses, 28(10), 1509–1518.

Kendon, A. (2017). Pragmatic functions of gestures: Some observations on the history of their study and their nature. Gesture, 16(2), 157-175.

Kısa, Y. D., Goldin-Meadow, S., & Casasanto, D. (2022). Do gestures really facilitate speech production? Journal of Experimental Psychology: General. 151(6), 1252–1271.

Kısa, Y. D. (2022). A communicative account of gesturing when speaking is difficult.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. Gesture, speech, and sign, 93–116.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. R package version, 1(1), 3.

Lucero, C., Zaharchuk, H., & Casasanto, D. (2014). Beat gestures facilitate speech production. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 36, No. 36).

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.

McNeill, D. (2008). Gesture and thought. University of Chicago press.

Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech. Language and gesture, 2, 186.

Ragsdale, J. D., & Fry Silvia, C. (1982). Distribution of kinesic hesitation phenomena in spontaneous speech. Language and Speech, 25(2), 185–190.

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. Team, R. C. (2020). R: A language and environment for statistical computing.

Seyfeddinipur, M., & Kita, S. (2001, June). Gestures and self-monitoring in speech production. In Annual Meeting of the Berkeley Linguistics Society (Vol. 27, No. 1, pp. 457-464).

Yap, D., Brookshire, G., & Casasanto, D. (2018). Beat gestures encode spatial semantics. In Proceedings of the 40th annual meeting of the cognitive science society.