AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2022, Vol. 151, No. 6, 1252–1271
https://doi.org/10.1037/xge0001135

Do Gestures Really Facilitate Speech Production?

Yağmur Deniz Kısa[1], Susan Goldin-Meadow[1], and Daniel Casasanto[2, 3]
[1] Department of Psychology, University of Chicago
[2] Department of Human Development, Cornell University
[3] Department of Psychology, Cornell University

Why do people gesture when they speak? According to one influential proposal, the Lexical Retrieval Hypothesis (LRH), gestures serve a cognitive function in speakers' minds by helping them find the right spatial words. Do gestures also help speakers find the right words when they talk about abstract concepts that are spatialized metaphorically? If so, then preventing people from gesturing should increase the rate of disfluencies during speech about both literal and metaphorical space. Here, we sought to conceptually replicate the finding that preventing speakers from gesturing increases disfluencies in speech with literal spatial content (e.g., the rocket went up), which has been interpreted as evidence for the LRH, and to extend this pattern to speech with metaphorical spatial content (e.g., my grades went up). Across three measures of speech disfluency (disfluency rate, speech rate, and rate of nonjuncture filled pauses), we found no difference in disfluency between speakers who were allowed to gesture freely and speakers who were not allowed to gesture, for any category of speech (literal spatial content, metaphorical spatial content, and no spatial content). This large dataset (7,969 phrases containing 2,075 disfluencies) provided no support for the idea that gestures help speakers find the right words, even for speech with literal spatial content. Upon reexamining studies cited as evidence for the LRH and related proposals over the past 5 decades, we conclude that there is, in fact, no reliable evidence that preventing gestures impairs speaking. Together, these findings challenge long-held beliefs about why people gesture when they speak.

Keywords: gesture, space, metaphor, speech production, Lexical Retrieval Hypothesis

Why do we gesture when we speak? For some gestures, it is self-evident that the speaker intends them to be communicative, as when we wave "hello" or give a "thumbs up" gesture to signal approval. Do speakers also gesture because gesturing serves a cognitive function in the speaker's mind, helping them to think or to talk?

For decades, researchers have posited that gestures facilitate speech production (Butterworth & Hadar, 1989; Krauss, 1998; Krauss & Hadar, 1999). According to an influential version of this proposal, the Lexical Retrieval Hypothesis (LRH), gestures facilitate speech production by helping speakers find the right words; however, only *some* gestures are posited to affect *some* words. Krauss and colleagues (Krauss, 1998; Rauscher et al., 1996) noted that people gesture far more frequently during phrases with spatial content than during phrases without it, and hypothesized that *gestures that reflect spatial features of meaning* helps speakers find the right *spatial words*. Krauss and colleagues hypothesized that some gestures "derive from knowledge encoded in a spatial format" (Rauscher et al., 1996, p. 227); the spatial features of a gesture (e.g., upward trajectory) facilitate production of words by priming the spatial features (e.g., upwardness) that enter into the search for that word (e.g., for the word "up"; Krauss & Hadar, 1999).[1]

What is the evidence that gesturing helps speakers find the right spatial words? According to Krauss and colleagues (Rauscher et al., 1996), if people have difficulty finding words, their speech should be more disfluent. If gesturing helps speakers find the right words, then preventing speakers from gesturing should make their speech production more disfluent, compared with when they are free to gesture. In an influential study, Krauss and colleagues reported that preventing people from gesturing increased the number of disfluencies they produced (pauses, repairs, etc.), and slowed down their speech selectively for the production of spatial phrases (Rauscher et al., 1996). Additionally, preventing people from gesturing when they produced spatial

This article was published Online First December 2, 2021.
Daniel Casasanto https://orcid.org/0000-0002-2021-1580
This research was supported in part by a James S. McDonnell Foundation Scholar Award (220020236) and a National Science Foundation Grant (BCS 125710) to Daniel Casasanto. Thanks to Kyle Jasmin, Cordelia Achen, and Corinne Roma for help with data coding, and to Howard Nusbaum and the members of the Goldin-Meadow lab and the Experience and Cognition Lab for discussion. This study was included in Yağmur Deniz Kısa's Master's thesis at the University of Chicago. A preliminary report on these results was presented at the 32nd Annual Convention of the Association for Psychological Science. Data are available on the Open Science Framework (https://osf.io/deurf/).
Correspondence concerning this article should be addressed to Daniel Casasanto, Department of Psychology, Cornell University, Martha Van Rensselaer Hall, Ithaca, NY 14853, United States. Email: casasanto@alum.mit.edu

[1] The LRH proposes that any gesture that expresses spatial semantics should help with the retrieval of spatial words. Previous research has shown that some beat gestures express spatial semantics (Yap et al., 2018). Therefore, given the proposed mechanism of the LRH, beat gestures that express spatial semantics should help with the retrieval of spatial words.

clauses increased their rate of filled pauses within those clauses (nonjuncture filled pauses), relative to filled pauses at the junctures of the clauses (i.e., juncture filled pauses), which Rauscher and colleagues claimed is the measure that most sensitively reflects problems in word finding. This study is frequently cited as evidence that gesturing helps speakers find the right spatial words and is discussed in reviews summarizing the state of knowledge about speaker-internal functions of gesture (e.g., Goldin-Meadow, 1999; Hostetter & Alibali, 2008). Largely based on this one study, the field has come to accept that one speaker-internal cognitive function played by gesture is to help speakers find the right words and facilitate speech production (e.g., Alibali et al., 2000, 2011; Casasanto, 2013; Goldin-Meadow, 1999; Hoetjes et al., 2014; Hostetter, 2011; Iverson & Goldin-Meadow, 1998; Krauss, 1998).

If LRH only explains how gestures help people talk about *space*, however, it provides only a limited account of how gesturing helps speaking. People spend a lot of time speaking about nonspatial ideas, including highly abstract concepts: entities like *time* and *value* that have no spatial magnitude, direction, or location. Yet, there is abundant evidence that people use space metaphorically to speak, think, and gesture about abstract concepts (Cienki, 2005; Lakoff & Johnson, 1980; McNeill, 1992; for a review, see Casasanto & Bottini, 2014). In one study, on which the present study builds, people spontaneously produced gestures whose form reflected the spatial direction implied in their speech (e.g., upward), regardless of whether they talked about concrete space (e.g., "the rocket *went up*") or metaphoric space (e.g., "my grades *went up*"; Yap et al., 2018). People's gestures reflected the predicted spatial directions (e.g., *better* is metaphorically upward), even when they talked about abstract concepts without using any spatial words (e.g., "my grades *got better*"). According to this study, which analyzed over 5,000 gestures, people were just as likely to gesture spontaneously in the predicted directions for metaphorically spatialized concepts (e.g., grades are not the kind of entity that can literally rise in space) as for literal spatial concepts. These results suggest that, like words for literal spatial concepts, words for metaphorical spatial concepts correspond to particular kinds of spatial information in speakers' minds (e.g., schematic representations of upward, downward, rightward, or leftward space), even when these words have no literal spatial uses (e.g., the word "better" cannot be used sensibly to denote literal spatial locations or paths).

If spatial information is activated in memory not only when people produce literal spatial language, but also when they produce metaphorical spatial language, then the same gestural mechanism should help people find words for both literal spatial scenarios and metaphorically spatialized ideas. Gesturing upward, for example, should help speakers not only to produce words or phrases like "my rocket went up," but also to produce words or phrases like "my grades went up" and perhaps even "my grades got better." If so, this discovery would substantially expand the scope of gesture's role in speech production, particularly since spatial schemas appear to be part of people's mental representations in many nonspatial conceptual domains that become spatialized metaphorically in language and thought, including time (Clark, 1973b), number (Dehaene et al., 1993), emotional valence (Casasanto, 2009), power (Schubert, 2005), similarity (Casasanto, 2008), intimacy (Matthews & Matlock, 2011), and musical pitch (Rusconi et al., 2006), among others (Lakoff & Johnson, 1980).

Here, we tested whether gestures serve a speaker-internal cognitive function by helping people find the right words with literal or metaphorical spatial content. We sought first to conceptually replicate Rauscher and colleagues' (Rauscher et al., 1996) study testing whether gesturing helps speakers produce words for literal spatial scenarios, and then to determine whether this benefit extends to producing speech about metaphorically spatialized ideas. We compared how fluently people spoke when they were allowed to gesture freely and when they were prevented from gesturing as they told stories with either literal or metaphorical spatial content. To measure disfluency in speech production, we calculated the number of speech disfluencies (repair, repeat, etc.) per word, speech rate (number of words per minute), and the relative frequency of nonjuncture filled pauses, following Rauscher et al. (1996). If gesturing only helps people find the right concrete spatial words, as suggested by Rauscher and colleagues (Rauscher et al., 1996), then preventing people from gesturing should make speech production more disfluent (i.e., higher disfluency rate, slower speech rate, and more nonjuncture than juncture filled pauses) only for speech with literal spatial content. Alternatively, if gesturing can also help speakers find the right abstract words, then preventing people from gesturing should make speech production more disfluent for speech not only with literal spatial content, but also with metaphorical spatial content.

Several studies in the 5 decades before and after Rauscher et al. (1996) have also tested whether preventing speakers from gesturing make their speech production more disfluent and reported null effects (Cravotta et al., 2018; Finlayson et al., 2003; Graham & Heywood, 1975; Hoetjes et al., 2014; Hostetter et al, 2007; Rimé et al., 1984). However, these null effects could simply be because none of the studies distinguished effects of gesture prevention on spatial and nonspatial speech; thus, arguably, these studies did not attempt to conceptually replicate Rauscher et al.'s (1996) claim that gesture prevention selectively affects spatial speech. Here, we tested this claim explicitly.

To preview our findings, contrary to our expectations based on earlier claims, preventing gesture had no significant effect on any of our planned dependent measures—not for speech with metaphorical spatial content, and not even for speech with literal spatial content. Due to the large number of data points in our study, it is not likely that the absence of these effects was the result of low statistical power. In response to this unexpected outcome, we first scrutinized our own data to confirm that there was no effect of gesture prevention *beyond* our planned analyses, in any subsets of the data (i.e., disfluency rates for different types of disfluencies). We then scrutinized the results of previous studies testing the LRH and related proposals, over the past 5 decades. Upon reexamining these studies, we conclude that (as in the present study), preventing gesture had no interpretable effect on speech production, even for literal spatial words, motivating a reexamination of widely held beliefs about why people gesture when they speak.

Method

Participants

Fifty-six Stanford University undergraduates (28 male) were recruited in pairs, and participated for course credit after giving informed consent (the study was approved by Stanford University's Institutional Review Board).

Materials

There were 12 brief stories in total, each 50–100 words, implying motion or extension in one of two spatial axes: horizonal or vertical (see online supplemental materials for an example story transcript for each of the 12 stories). Four of the stories had literal spatial content, describing actual spatial scenarios in the physical world using concrete spatial directions (e.g., "the rocket went *higher*"; "the scuba diver went *down*"). Eight of the stories had metaphorical spatial content, describing abstract nonspatial phenomena that are nevertheless commonly talked and thought about using spatial directions metaphorically (e.g., "my grades went *higher*"; "the price went *down*"). Each of the eight metaphorical stories had two versions: metaphorical stories with spatial language and metaphorical stories without spatial language. Metaphorical stories with spatial language described nonspatial phenomena using spatial words or phrases in their abstract metaphoric senses (e.g., "my grades went *higher*"). Metaphorical stories *without* spatial language were identical to metaphorical stories with spatial language, except that spatial words or phrases that are used metaphorically were replaced with nonspatial paraphrases conveying nearly the same meaning and implying the same spatial directions (e.g., "my grades got *better*"). The stories involved an overall spatial direction (horizontal or vertical); however, they contained words and phrases that expressed spatial ideas other than directions or positions such as "long," "crossing," "around," "plunge," "stretch back," "boost up," "stuck at the top," and so forth.

Procedure

Participants were told that the experiment was about storytelling. They took turns studying written stories, each for 60 s, and then retelling the stories to their partners. They 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.

After starting with a warm-up story, each participant retold six stories in randomized order: two stories with literal spatial content and four stories with metaphorical spatial content. Each pair of participants received only one version of each metaphorical story: either with spatial language or without spatial language (i.e., one pair of participants would receive either the story about "grades going higher" or the story about "grades getting better").

Each pair of participants was assigned to one of two gesture conditions: gesture prevented or gesture allowed. In the gesture prevented condition, participants were instructed to hold down keys on a computer keyboard, one key with each hand, during the entire time they were retelling the stories. They were told that the keys activated the microphones mounted on top of the computer monitor in front of them; in fact, the microphones were nonfunctional. In the gesture allowed condition, participants simply told the stories without being instructed to hold down keys on a keyboard; they were not told to gesture. Testing lasted 20–30 min.

Coding

Analyses of the gestures from the Gesture Allowed condition were reported in Yap et al. (2018), but no analysis of speech disfluencies was reported, and no data from the Gesture Prevented condition have been reported previously. In the Gesture Allowed condition, participants produced a total of 2,249 gestures including 1,609 beats, 328 iconic gestures, 252 deictics, 48 metaphoric gestures, 10 adaptors, and two emblems. Beats were categorized solely based on form, following McNeill's (1992) *beat* filter. When categorized based on meaning with respect to accompanying speech, 629 of the 1,609 beats reflected the spatial ideas expressed in the accompanying speech. Furthermore, the overall rate of gestures that would be predicted to facilitate spatial speech (i.e., beats reflecting spatial semantics, iconic gestures, deictics, and metaphoric gestures) in the Gestures Allowed condition was 56% (1,257 out of 2,249 gestures). We used ELAN (Wittenburg et al., 2006) to code speech disfluencies.

Speech Content Coding

Participants' audio recordings of the stories were transcribed verbatim; 22 of the 336 stories were excluded because the speech was inaudible. The transcriptions of participants' audio recordings of the stories were parsed into clauses and phrases. Coder 1 determined whether each phrase had spatial content, literal or metaphorical. A given phrase was classified as having spatial content if it contained language that implies literal or metaphorical motion, extent, or position along either the lateral or vertical axis. For instance, "went higher" in the rocket story would be a phrase with spatial content since the overall story has a vertical spatial schema of a rocket going higher and higher, and "higher" implies literal motion along the vertical axis. Similarly, "came back down" in the rocket story would be a phrase with spatial content, because "down" also implies literal motion along the vertical axis. Alternatively, phrases were classified as having no spatial content if they did not imply a literal or metaphorical spatial schema.

Spatial Content Type Coding

Phrases with spatial content were classified as literal or metaphorical, and phrases with metaphorical spatial content were further classified as having or not having spatial language, using the same criteria used to construct the stories.

Participants produced a total of 7,969 spoken phrases, with very similar numbers of phrases across gesture conditions (Gesture allowed 4,000; Gesture prevented: 3,969). Overall, 2,801 phrases (35% of all phrases) included spatial content, with 962 literal and 1,839 metaphorical spatial content. Coder 2 determined speech content for 10% of all phrases and the intercoder agreement for speech content was 96% (Cohen's $\kappa = .92$, $z = 26.1$, $p < .0001$).

Speech Disfluency Coding

Disfluency Rate Coding. Coder 1 recorded the location and type of speech disfluency for each story. Speech disfluencies included repeats, repairs, filled pauses (uh, um, etc.), and unfilled pauses. Coder 2 coded the number of speech disfluencies for a random subset of the stories (29 stories out of a total of 314 stories: about 10%). The intercoder agreement for the number of speech disfluencies for each story was very high; Coder 2's coding explained 94% of the variance in Coder 1's coding (intercoder correlation for number of disfluencies per story: $\beta = .94$, $R^2 = .94$, $p < .0001$).

Participants produced a total of 2,075 disfluencies (Gesture allowed: 1,041; Gesture prevented: 1,034). Speech disfluencies included 905 (44%) filled pauses, 492 (24%) repairs, 446 (21%) unfilled pauses, and 232 (11%) repeats. For each story, we calculated the total number of disfluencies that occurred during phrases with spatial content (both overall and for each spatial content type separately) and during phrases without spatial content. Finally, for each story, we calculated the total number of words in phrases with and without spatial content, which we used as a baseline in our analysis.

The analyses on disfluency rate were done at the story level, rather than at the phrase level, to reduce the inflated zero problem of our count data (i.e., too many phrases with no disfluencies). When comparing the overall effect of gesture prevention on disfluency rate, we had one observed rate for each story; when comparing the effect of gesture prevention for speech with spatial content versus for speech without spatial content, we had two observed rates for each story; and so on.

Speech Rate Coding. For each phrase, we measured its duration and coded the number of words it contained. We computed the speech rate for each phrase, dividing the number of words by the duration for each phrase. Participants had an average speech rate of 212 words per minute at the phrase level (SD = 86.25). When we computed the speech rate for each story (N = 336), dividing the total number of words by the total duration of all the phrases in each story, participants had an average speech rate of 171 words per minute (SD = 38.86).[2] The analyses were done with speech rates calculated at the phrase level.

Nonjuncture Filled Pause Coding. Each filled pause (905 in total) was classified as nonjuncture or juncture, based on whether it occurred between clause junctions, that is, within a clause (e.g., "And the rocket *um* shoots up as high as the 10th floor"), or at clause junctures (e.g., "*Um* and the rocket shoots up as high as the 10th floor"). Twelve filled pauses were excluded from the analyses since they occurred during clauses with mixed spatial content type (e.g., both literal and metaphorical spatial content).

Analysis

We conducted all analyses by fitting generalized linear mixed-effect models, using R (R Core Team, 2020), the glmer() function in the lme4 library (Bates et al., 2015), and the optimx package (Nash & Varadhan, 2011). We treated Subject (N = 56) and Story (N = 12) as random effects, including random intercepts for both in analyses, since our outcome variable, disfluency, is likely to vary across different subjects and stories. Subjects and items (i.e., stories here) usually do vary idiosyncratically not only in their global mean responses, but also in their sensitivity to experimental treatment. We used "maximal" random effect structures justified by our design (Barr et al., 2013), including not only random intercepts for Subject and Story, but also random slopes for our fixed factors that are within-subject or within-story, allowing disfluencies of subjects and stories to vary differentially based on our fixed factors (e.g., subjects could be affected differently by speech content manipulation in their production of disfluencies).

Gesture condition (gesture allowed; gesture prevented) was a between-subjects and within-story (i.e., within-item) factor. Speech content (speech with spatial content; speech without spatial content) was a within-subject and within-story factor. Spatial content type (literal; metaphorical with spatial language; metaphorical without spatial language) was a within-subject and between-story factor. When testing for interactions, we included random slopes only for the highest-order combination of within-unit factors subsumed by each interaction (Barr, 2013; e.g., by-story random slope for Gesture Condition × Speech Content interaction). We used likelihood ratio tests (LRTs) to test for fixed effects, with post hoc contrasts performed on subsets of the data. When the models failed to converge, we simplified the random effects structure of the maximal model by (a) dropping the correlation between the random intercept and random interaction slope, (b) dropping the intercepts (Barr et al., 2013), (c) dropping the random interaction slope, and (d) dropping any random slope (i.e., intercept only model).

Results

Disfluency Rate

To compare speech disfluency rate across experimental conditions, we used mixed-effects Poisson regressions. We incorporated the number of words as an offset term into the model so that we modeled speech disfluency rate (number of speech disfluencies per word), rather than raw count data of number of speech disfluencies (Agresti, 2003). Overall, Poisson regression with an offset term allowed us to model rate data with successfully approximating a normal distribution and constant variance of residuals.

Overall Effect of Gesture Prevention on Disfluency Rate

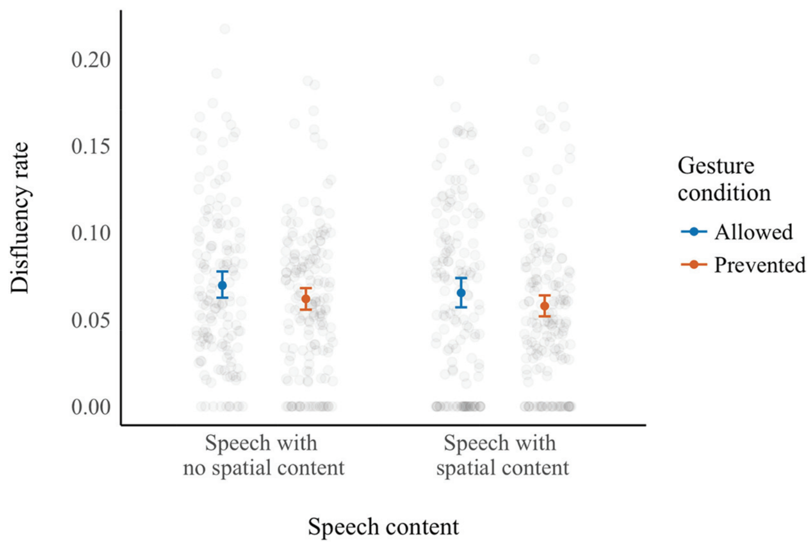
Did people produce a higher rate of disfluencies when they were prevented from gesturing, compared with when they were allowed to gesture? In a first analysis including all speech content types (with and without spatial content), we found no evidence of an effect of gesture prevention on rate of speech disfluencies. Disfluency rates when people were prevented from gesturing (M = .07, SD = .04, *Median* = .07) were statistically indistinguishable from disfluency rates when people were allowed to gesture (M = .06, SD = .03, *Median* = .06; $\chi^2(1) = 1.07$, $p = .30$).

Effect of Gesture Prevention During Speech With Versus Without Spatial Content

A second analysis tested the effect of gesture prevention on disfluency rates in speech with spatial content and in speech with no spatial content. Results showed that preventing people from gesturing had no significant effect on disfluency rates during speech with spatial content ($\chi^2(1) = 1.71$, $p = .19$) or during speech with no spatial content ($\chi^2(1) = .56$, $p = .45$). Notably, the nonsignificant trends went in the opposite direction of what the LRH would predict: People were slightly *less* disfluent when prevented from gesturing compared with when they are allowed to gesture, both during speech with spatial content and during speech with no spatial content (see Figure 1). The (non-)effect of gesture prevention did not differ significantly between speech with and without spatial content, as indicated by a

[2] Calculating speech rate at the phrase level results in an average speech rate that is faster compared with when we calculate speech rate at the story level, simply by virtue of the different order of operations involved in each of these ratio calculations. Calculating speech rate at the phrase level involves first dividing the number of words for each phrase by the duration of each phrase and only then taking the average. On the other hand, calculating speech rate at the story level involves first adding the number of words of all the phrases in a story and adding the duration of all the phrases in a story, then dividing the total number of words in a story by the total duration of phrases in a story and only then taking the average.

Figure 1
Disfluency Rates During Speech With Spatial Content (Right) and Speech With No Spatial Content (Left)



Note. Gray dots show individual data points. Error bars show bootstrapped 95% confidence interval (CI) around the group means. See the online article for the color version of this figure.

nonsignificant interaction between Gesture condition and Speech content ($\chi^2(1) = .30$, $p = .58$).

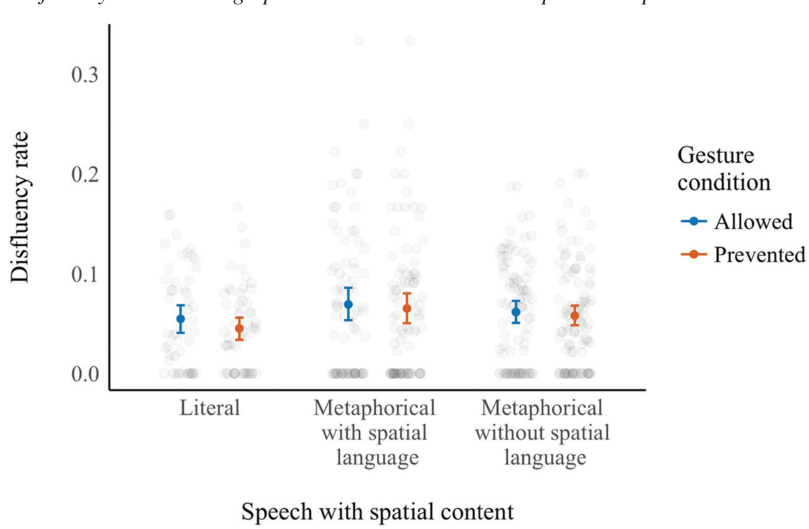
Effect of Gesture Prevention for Literal Versus Metaphorical Spatial Content

A third set of analyses tested for effects of preventing gesture during speech with literal spatial content and with metaphorical spatial content. Results showed no significant effect of gesture prevention on disfluency rate for any type of spatial content (Literal: $\chi^2(1) = .97$, $p = .32$; Metaphorical with spatial language: $\chi^2(1) = 1.56$, $p = .21$; Metaphorical without spatial language: $\chi^2(1) = 1.09$, $p = .30$), and the (non-)effect of gesture prevention on disfluency rate did not differ across these conditions ($\chi^2(2) = .07$, $p = .97$; see Figure 2).

Effect of Gesture Prevention for Different Disfluency Types

Did people produce a higher rate of any kind of disfluency (i.e., repairs, repeats, filled pauses, or unfilled pauses) when they were prevented from gesturing, compared with when they were allowed to gesture? There was no effect of preventing gesture on the rate of any type of disfluency: repairs ($\chi^2(1) = .04$, $p = .85$),

Figure 2
Disfluency Rates During Speech With Literal and Metaphorical Spatial Content



Note. Gray dots show individual data points. Error bars show bootstrapped 95% confidence interval (CI) around the group means. See the online article for the color version of this figure.

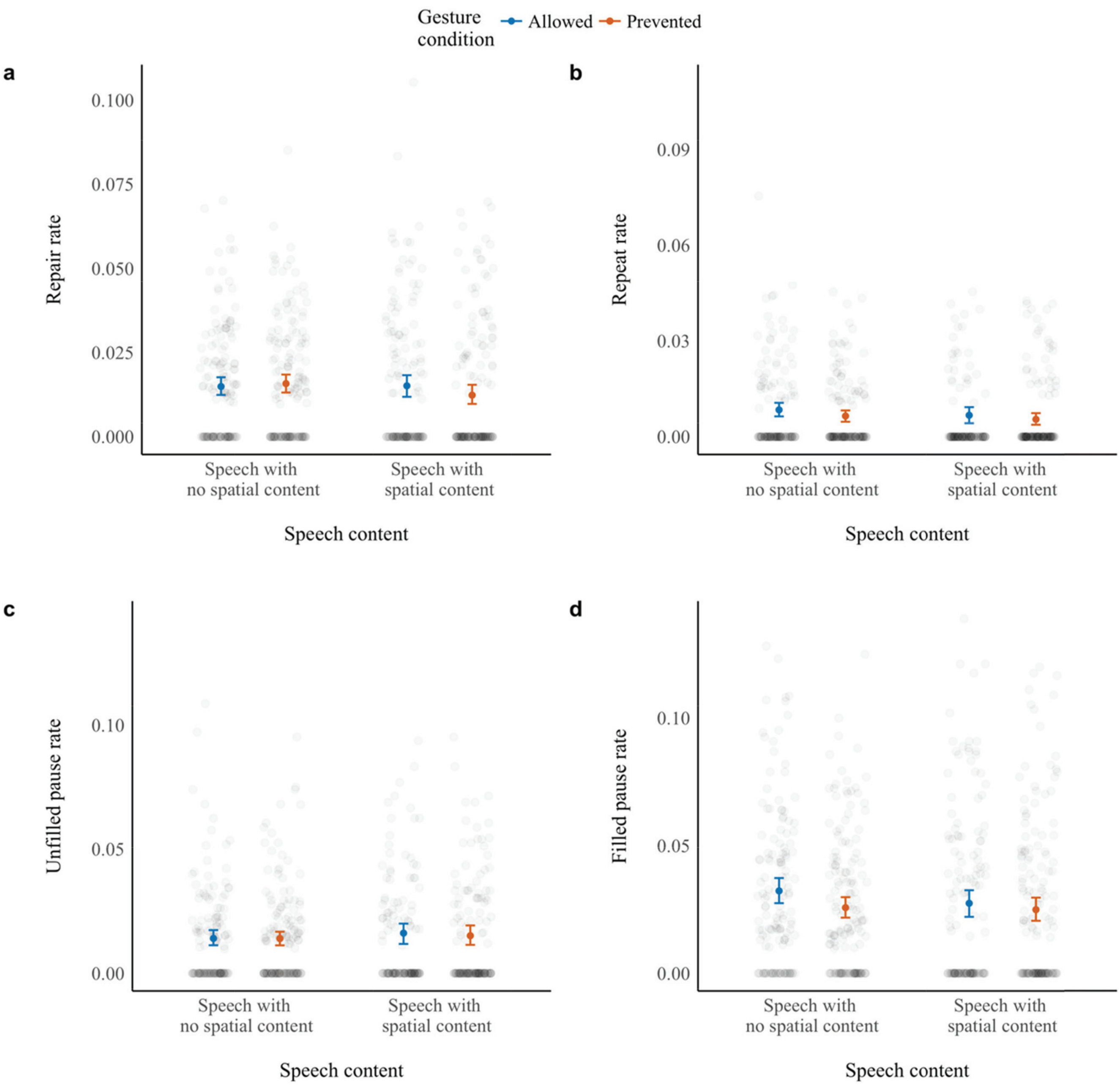
repeats ($\chi^2(1) = 1.01$, $p = .31$), unfilled pauses ($\chi^2(1) = .02$, $p = .89$), or filled pauses ($\chi^2(1) = 1.05$, $p = .30$). Similarly, there was no effect of preventing gesture on disfluency rates in speech with spatial content across different disfluency types: repairs ($\chi^2(1) = .54$, $p = .46$), repeats ($\chi^2(1) = .49$, $p = .48$, unfilled pauses ($\chi^2(1) = .11$, $p = .74$), or filled pauses ($\chi^2(1) = .85$, $p = .35$). Notably, the nonsignificant trends went in the opposite direction of what the LRH would predict: For any kind of disfluency, people were slightly less disfluent when prevented from gesturing compared with when they are allowed to gesture during speech with spatial content (see Figure 3).

Speech Rate

To compare speech rate for each phrase across experimental conditions, we used mixed-effects Gaussian regressions.

Figure 3

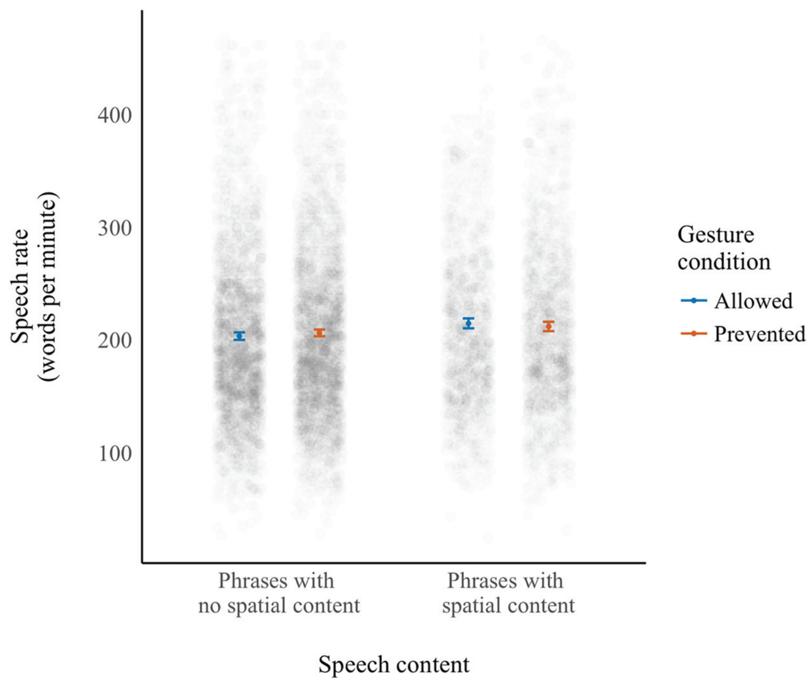
Disfluency Rates During Speech With Spatial Content and Speech With No Spatial Content Across Different Types of Disfluency: (a) Repairs, (b) Repeats, (c) Unfilled Pauses, and (d) Filled Pauses



Note. Gray dots show individual data points. Error bars show bootstrapped 95% confidence interval (CI) around the group means. See the online article for the color version of this figure.

Figure 4

Speech Rates (In Words Per Minute) During Phrases With Spatial Content (Right) and Phrases With No Spatial Content (Left)



Note. Gray dots show individual data points. Outliers (data points that are 3 *SD*s above the mean) are removed for visualization purposes. Error bars show bootstrapped 95% confidence interval (CI) around the group means. See the online article for the color version of this figure.

Overall Effect of Gesture Prevention on Speech Rate

Did people produce slower speech when they were prevented from gesturing, compared with when they were allowed to gesture? In a first analysis including all speech content types (with and without spatial content), we found no evidence of an effect of preventing gesture on speech rate. Speech rates when people were prevented from gesturing (M = 212.98, SD = 86.91) were statistically indistinguishable from speech rates when people were allowed to gesture (M = 211.25, SD = 85.53, $\chi^2(1)$ = .08, p = .77).

Effect of Gesture Prevention During Phrases With Versus Without Spatial Content

We tested the effect of preventing gesture on speech rate in phrases with spatial content and in phrases with no spatial content. Results showed that preventing people from gesturing had no significant effect on people's speech rates during phrases with spatial content ($\chi^2(1)$ = .001, p = .97) or during phrases with no spatial content ($\chi^2(1)$ = .23, p = .63). The (non-)effect of gesture prevention did not differ significantly between phrases with and without spatial content, as indicated by a nonsignificant interaction between Gesture condition and Speech content ($\chi^2(1)$ = 1.36, p = .24; see Figure 4)[3].

Effect of Gesture Prevention for Literal Versus Metaphorical Spatial Content

We tested for effects preventing gesture on speech rate during speech with literal spatial content and with metaphorical spatial content. Results showed no significant effect of preventing gesture on speech rate for any type of spatial content (Literal: $\chi^2(1)$ = .10, p = .75; Metaphorical with spatial language: $\chi^2(1)$ = .42, p = .52; Metaphorical without spatial language: $\chi^2(1)$ = .11, p = .74), and the (non-)effect of gesture prevention on speech rate did not differ across these conditions ($\chi^2(2)$ = 1.16, p = .56; see Figure 5).

Rate of Nonjuncture Filled Pauses

To compare nonjuncture filled pause rate across experimental conditions, we used mixed effects logistic regressions with nonjuncture or juncture as the binary outcomes for a filled pause.

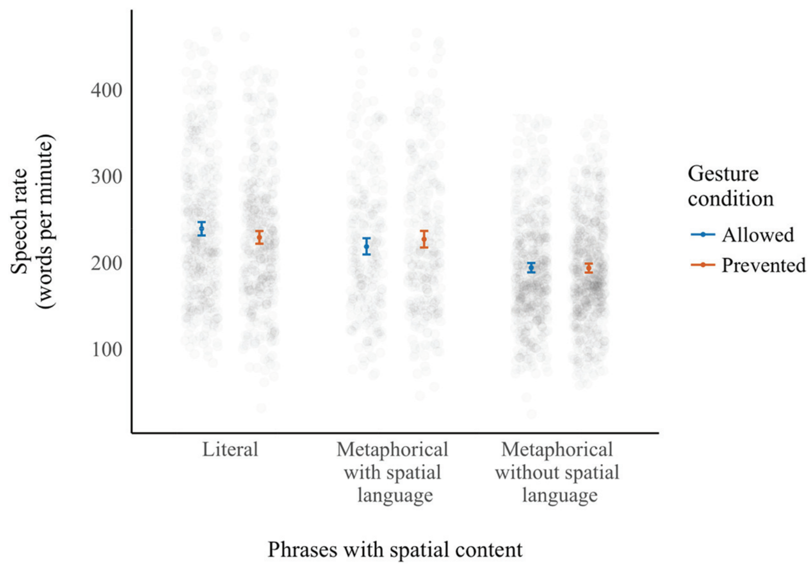
Overall Effect of Gesture Prevention on the Rate of Nonjuncture Filled Pauses

Did people produce a higher rate of nonjuncture filled pauses when they were prevented from gesturing, compared with when they were allowed to gesture? In a first analysis including all speech

[3] The way we calculated speech rates here and in the analyses is different from Rauscher et al. (1996): We calculated speech rates for each phrase, whereas Rauscher et al. (1996) calculated speech rates for each participant. Calculating speech rate for each phrase results in an average speech rate that is faster than if we were to calculate speech rate for chunks of speech that are larger than a phrase, simply by virtue of the different order of operations (addition vs. division) involved in each of these ratio calculations. Therefore, a direct comparison between the speech rates across the two studies is difficult. However, we can compare speech rates between the experimental conditions within each study.

Figure 5

Speech Rates (In Words Per Minute) During Phrases With Literal and Metaphorical Spatial Content



Note. Gray dots show individual data points. Outliers (data points that are 3 *SD*s above the mean) are removed for visualization purposes. Error bars show bootstrapped 95% confidence interval (CI) around the group means. See the online article for the color version of this figure.

content types (with and without spatial content), we found no evidence of an effect of preventing gesture on the rate of nonjuncture filled pauses. The nonjuncture filled pause rate when people were prevented from gesturing (184 nonjuncture filled pauses out of 434 filled pauses) was statistically indistinguishable from the same rate when people were allowed to gesture (216 nonjuncture filled pauses out of 459 filled pauses, $\chi^2(1) = .87$, $p = .35$; see Table 1).

Effect of Gesture Prevention During Phrases With Versus Without Spatial Content

We tested the effect of preventing gesture on nonjuncture filled pause rate in phrases with spatial content and in phrases with no spatial content. We found that preventing people from gesturing had no significant effect on people's nonjuncture filled pause rate during phrases with spatial content ($\chi^2(1) = .20$, $p = .65$) or during phrases with no spatial content ($\chi^2(1) = .47$, $p = .49$). The (non-) effect of gesture prevention did not differ significantly between phrases with and without spatial content, as indicated by a nonsignificant interaction between Gesture condition and Speech content ($\chi^2(1) = .29$, $p = .59$; see Figure 6).

Table 1

Number of Nonjuncture Filled Pauses (Out of the Total Number of Filled Pauses) During the Gestures Allowed and Gestures Prevented Conditions

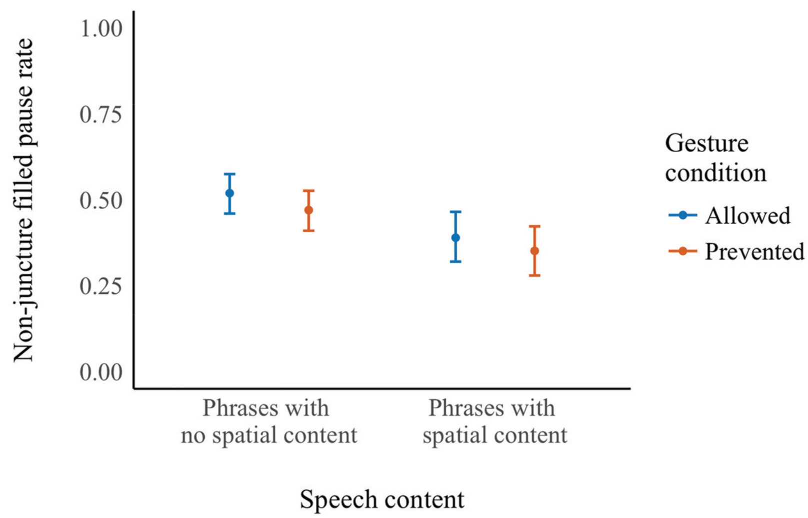
Speech content	Gesture condition	
	Gestures allowed	Gestures prevented
Phrases with no spatial content	149 (287)	125 (266)
Phrases with spatial content		
Literal	23 (48)	9 (33)
Metaphorical with spatial language	12 (34)	11 (29)
Metaphorical without spatial language	32 (90)	39 (106)
All phrases	216 (459)	184 (434)

Effect of Gesture Prevention for Literal Versus Metaphorical Spatial Content

We tested for effects of preventing gesture on nonjuncture filled pause rate during speech with literal spatial content and with metaphorical spatial content. The effect of gesture prevention on nonjuncture filled pause rate was not significant for phrases with metaphorical spatial content (Metaphorical with spatial language: $\chi^2(1) = .06$, $p = .81$; Metaphorical without spatial language: $\chi^2(1) = .05$, $p = .82$). This effect was nominally significant (i.e., $p < .05$) for phrases with literal spatial content (Literal: $\chi^2(1) = 4.13$, $p = .04$); however, the effect went in the opposite direction from what the LRH would predict: People produced a *lower* rate of nonjuncture filled pauses when prevented from gesturing compared with when they are allowed to gesture. We do not interpret this backward result, for two reasons. First, this result is "significant" only if we maintain a nonconservative alpha value (.05) that is not corrected for multiple comparisons. Alpha correction would be necessary, especially for interpreting a single nonpredicted result, given that our search for *any* significant effect of gesture prevention led us to conduct 34 independent tests (22 planned, 12 post hoc); after appropriate correction, this nonpredicted result is no longer statistically significant. Second, this nonpredicted effect for phrases with literal spatial content did not differ significantly from the (non-)effects of gesture prevention for phrases with metaphorical spatial content, as indicated by a nonsignificant interaction between Gesture condition and Spatial content type ($\chi^2(2) = 2.92$, $p = .23$; see Figure 7).

Figure 6

Nonjuncture Filled Pause Rate (i.e., Proportion of Filled Pauses That Are Nonjuncture as Opposed to Juncture) During Phrases With Spatial Content (Right) and Phrases With No Spatial Content (Left)



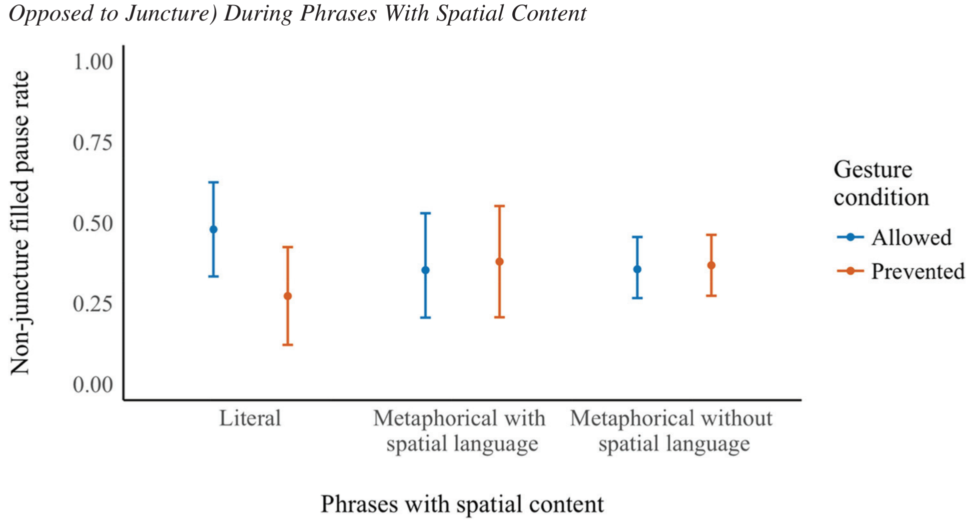
Note. Error bars show bootstrapped 95% confidence interval (CI) around the group means. For exact counts of filled pauses across conditions see Table 1. See the online article for the color version of this figure.

Discussion

Does gesturing facilitate speech production by helping people find the right spatial words? We found no evidence that speakers speak less fluently when they are prevented from gesturing, compared with when they are allowed to gesture freely. Preventing gesture did not have the predicted effect on disfluency rate, speech rate, or nonjuncture filled pause rate overall, or on any of the experimental conditions we evaluated. We failed to find an effect of preventing gesture during speech with metaphorical spatial content, a finding that would have expanded the scope of LRH to encompass speech about abstract concepts. More fundamentally, we also found no significant effect of preventing gesture during speech with literal spatial content. We failed to find support for Rauscher and colleagues' (Rauscher et al., 1996) influential claim that preventing gesture increases disfluencies for spatial language. More broadly, our data provide no support for the idea that gesturing facilitates speech production by helping people find the right words.

Figure 7

Nonjuncture Filled Pause Rate (i.e., Proportion of Filled Pauses That Are Nonjuncture as Opposed to Juncture) During Phrases With Spatial Content



Note. Error bars show bootstrapped 95% confidence interval (CI) around the group means. For exact counts of filled pauses across conditions see Table 1. See the online article for the color version of this figure.

Why Did We Find No Evidence That Gestures Help Speakers Find the Right Spatial Words?

Why did preventing gestures have no effect on our participants' speech, even when speakers were using words like "up" and "down" to describe concrete spatial scenarios? A first possible explanation to consider for any null result may be lack of statistical power. This explanation is unlikely for our study, however, given (a) the size of our data set, (b) our analysis choices that leverage the large number of within-subject observations, and (c) the qualitative patterns in the data. Overall, we analyzed 7,969 phrases containing 2,075 disfluencies, resulting in a large number of observations per subject (e.g., 156 observations per subject in the speech rate analysis). Unlike Rauscher and colleagues' (Rauscher et al., 1996), we did not average across observations for each subject, but instead conducted our statistical analyses for each story, clause, or phrase the participants produced. Modeling all data points increased our statistical power by increasing within-subject observations, while also reducing a source of Type I error (i.e., false positives) that was present in Rauscher et al.'s (1996) analyses by accounting for item-wise variance (Baayen et al., 2008; Clark, 1973a). Finally, if a lack of power were responsible for our null effects, then, overall, we would expect to find trends in the predicted direction that failed to reach statistical significance. This was not the case. On the contrary, as noted above, the strongest trends went opposite from the predicted direction, including the only trend among our 34 comparisons that was "significant" at $p < .05$ before correction for multiple comparisons. Considering all of the between-condition comparisons shown in Figures 1–7 (i.e., comparing each pair of adjacent light gray and dark gray bars), about twice as many comparisons trended against the LRH as trended in support of it.

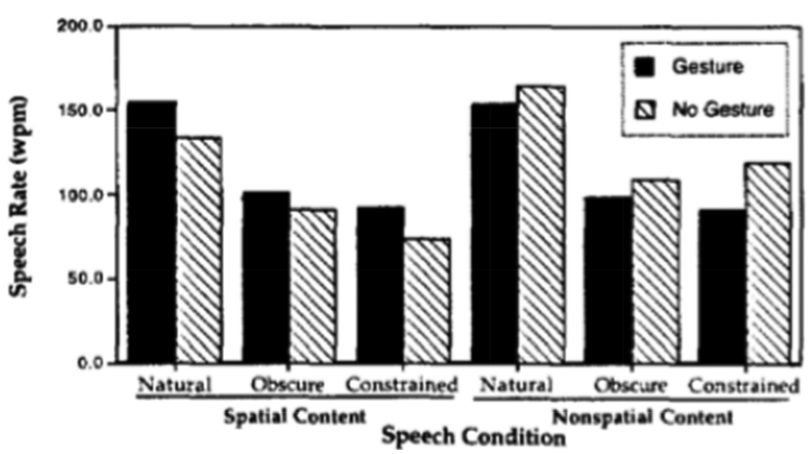
A second possible explanation for a null effect could rest in having the predicted effects "hidden" in subsets of the data and obscured by aggregating over conditions or trial types. To ensure that this was not the case, we report graphs and planned analyses for all subsets of the data, broken out not only by spatial versus nonspatial content, but also by multiple types of language (literal, metaphorical spatial, and nonspatial language). To ensure that effects of any particular type of disfluency were not being masked by noneffects for other types, we also conducted post hoc analyses of each disfluency type, individually (Figure 3); preventing gestures did not have the predicted effect on any type of disfluency.

A third possible explanation for our null results could rest in levels of difficulty in speech production. Might effects of gesture prevention on fluency only emerge during difficult production conditions—and if so, is it possible that our production task was simply too easy? An examination of Rauscher et al.'s (1996) results does not support this possible explanation. Rauscher and colleagues (Rauscher et al., 1996) built into their experimental design three different levels of difficulty producing speech, positing that more difficult conditions should increase disfluency. Their participants produced natural speech with normal demands on word production in one condition (normal-speech condition); in a second condition participants had to use as many obscure words as possible (obscure-speech condition); in a third condition they had to avoid using words that contained a specified letter (constrained-speech condition). The latter two conditions make higher demands on word production than the normal-speech condition, and produced higher disfluency rates, overall.

Did preventing gestures increase disfluencies more when speech production was more difficult? Rauscher and colleagues (Rauscher et al., 1996) did not report any analyses that addressed this question, but the trends shown in their plots do not support the possibility that greater production difficulty increases the effect of gesture prevention on disfluency (see Figures 8 and 9 below, reprinted from Rauscher et al., 1996). On the contrary, for nonjuncture filled pauses (Figure 9) the effects of gesture prevention in Rauscher et al. (1996) were numerically *smaller* when the speech production task was harder. Speech production in our task was clearly less challenging than in Rauscher et al.'s (1996) obscure speech and constrained speech conditions; it is hard to know

Figure 8

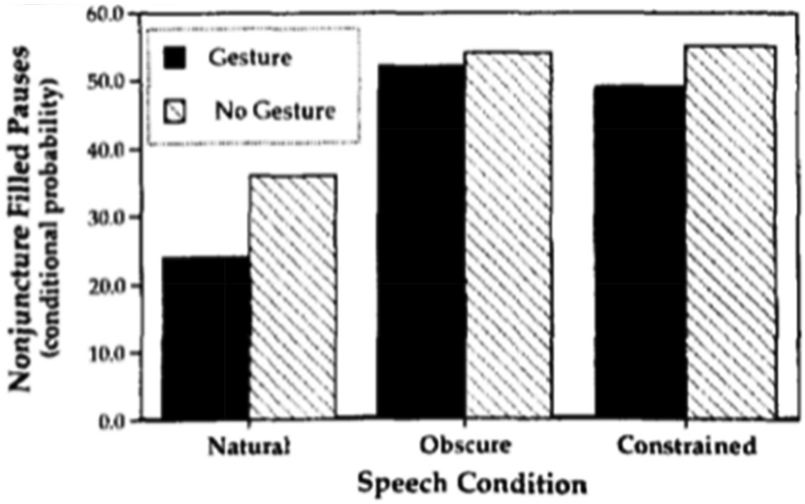
Speech Rate (Words Per Minute) in the Natural-, Obscure-, and Constrained-Speech Conditions for Speech With Spatial and Nonspatial Content and When Subjects Were and Were Not Allowed to Gesture



Note. Reprinted from "Gesture, Speech, and Lexical Access; The Role of Lexical Movements in Speech Production," by F. H. Rauscher, R. M. Krauss, and Y. Chen, 1996, *Psychological Science*, *7*(4), p. 229. Copyright [1996] by SAGE Publications. Reprinted with permission.

Figure 9

Probability of a Nonjuncture Filled Pause Given a Filled Pause in the Natural-, Obscure-, and Constrained- Speech Conditions for Speech With Spatial and Nonspatial Content and When Subjects Were and Were Not Allowed to Gesture



Note. Reprinted from "Gesture, Speech, and Lexical Access; The Role of Lexical Movements in Speech Production," by F. H. Rauscher, R. M. Krauss, and Y. Chen, 1996, *Psychological Science*, *7*(4), p. 229. Copyright [1996] by SAGE Publications. Reprinted with permission.

whether speech production in our task was more or less difficult than in Rauscher et al.'s "natural" speech condition, but Rauscher et al.'s results provide no evidence that making a speech production task more difficult yields a more sensitive test of effects of gesture prevention on disfluency.

Another possible explanation for our null results could rest in how complex the spatial ideas were that were expressed in speech. Might effects of gesture prevention on fluency only emerge during speech expressing complex spatial ideas—and if so, is it possible that our speech production task was semantically too simple? This explanation is unlikely to account for the null results in our study since the spatial descriptions in our stories were quite complex and varied involving position, shape, motion, and trajectory information (see online supplemental materials, e.g., story transcripts). The spatial words and phrases elicited in the current study cannot exhaust the range of complexity that could be present in a spatial description. Therefore, it is an open question whether gestures facilitate only highly complex spatial speech, beyond the level of complexity tested here.[4]

The most fruitful explanation for the null results in the present study, we believe, rests in a reexamination of Rauscher et al.'s (1996) results, and of other studies testing the LRH. We first reexamine Rauscher et al.'s (1996) study in extensive detail to evaluate their proposed evidence for the LRH, because this is the only study in the gesture literature that claims to have found evidence that gesture prevention makes spatial speech disfluent, a claim that we failed to replicate here, and is one of the most cited studies in the gesture literature that largely shaped theories of why people gesture when they speak. In aiming to answer the question whether gestures facilitate speech production, it is critical to reexamine this seminal study by Rauscher et al. (1996) that has been the major source of answer to this question since its publication. We then examine in detail other studies that aimed to test effects of gesture prevention on speech disfluency. Much to our surprise, a careful examination of the highly influential study by Rauscher et al. (1996) and other studies that tested effects on speech disfluency yields no clear evidence that preventing gesture increases disfluencies in speech.

Is There Any Evidence That Gestures Help Speakers Find the Right Spatial Words?

Rauscher and colleagues' (Rauscher et al., 1996) study has been widely cited as evidence for the idea that gesturing helps speakers find the right words and, more broadly, as some of the first evidence that gesturing serves a cognitive function for speakers (e.g., Alibali et al., 2000, 2011; Casasanto, 2013; Goldin-Meadow, 1999; Hoetjes et al., 2014; Hostetter, 2011; Iverson & Goldin-Meadow, 1998). Did Rauscher et al. (1996) find the pattern of results predicted by their hypothesis—that is, did their study show that gesturing helps speakers find the right spatial words? No. To support their conclusions, Rauscher and colleagues would need to have shown particular patterns of data in each of their three dependent measures (i.e., speech disfluency rate, speech rate, and nonjuncture filled pause rate). Below, we outline the predicted patterns for each of these

[4] See Table 2 for other studies that failed to find effects of gesture production on speech disfluency for speech with arguably higher levels of spatial complexity than found in our data, for example, describing how to tie a knot.

measures and explain why the observed patterns did not support these predictions.

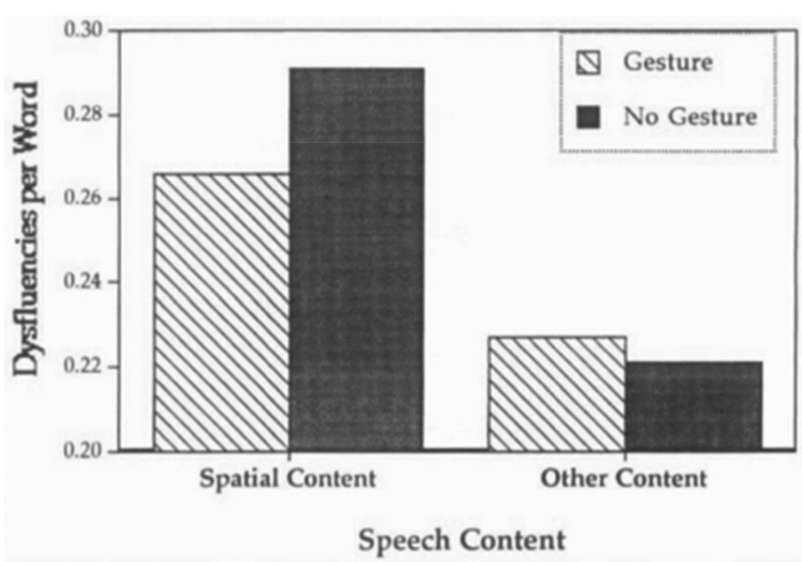
To preview these explanations, in each of the three dependent measures the LRH predicted (and required) two effects: First, preventing gesture should make spatial speech more disfluent, resulting in a simple effect of gesture prevention in speech with spatial content. Second, preventing gesture should increase disfluency *selectively* during speech with spatial content, as opposed to speech with no spatial content, resulting in a two-way interaction of gesture condition (gestures allowed; gestures prevented) and speech content (spatial content; no spatial content). The selectivity of this interaction effect is crucial for the data to support the LRH, which hypothesizes that there is a special link between gesture and *spatial* words. Furthermore, without this two-way interaction, the effect of gesture prevention would be open to a hypothesis-irrelevant interpretation: Preventing gesture on *any* kind of speech could increase disfluency simply due to the unnaturalness of preventing gesture (see Rauscher et al., 1996, p. 229 for a similar argument).

Speech Disfluency Rate

Rauscher et al.'s (1996) most influential claim is that preventing gesture causes higher disfluency rates *only* during speech with spatial content. Yet, the simple effect required to support this claim was not statistically significant, and the required two-way interaction was never reported, neither in Rauscher et al. (1996), nor in subsequent review articles and chapters highlighting these results (see Figure 4 in Krauss, 1998; reprinted here as Figure 10). Turning first to the required simple effect, the LRH predicted that participants should produce more disfluencies when they are prevented from gesturing, compared with when they are allowed to gesture, during speech with spatial content (i.e., the two bars on the left in Figure 10 should be significantly different from each other). However, this critical simple effect was only marginally significant, as reported (i.e., $p < .066$). Notably, even this reported value is anticonservative, in at least two ways. First, the statistical test did not account for item-wise variance (Baayen et al., 2008; Clark, 1973a), leading to an increased probability of Type I error (i.e., a false positive result). Second, the alpha value for pairwise comparisons was not corrected for the multiple statistical comparisons reported, and the even greater number of comparisons that could have been conducted in this $2 \times 2 \times 3$ design. After correcting the alpha value appropriately, the reported marginal p-value would no longer approach significance.

This simple effect was necessary to support Rauscher et al.'s (1996) main claim; because it was not significant, there is no need to analyze the speech disfluency rates results further to evaluate their significance. However, as noted, there was a second effect required by the LRH as well: the two-way interaction of gesture condition and speech content. The significance of this critical interaction was implied by Rauscher et al.'s (1996) text, but a *different* interaction was reported in its place: the three-way interaction of gesture condition (gesture allowed; gesture prevented), speech content (spatial content; nonspatial content), and speech condition (natural = producing natural speech; obscure = using as many obscure words as possible; constrained = avoiding using words that contained a specified letter). Although this three-way interaction was significant, it is irrelevant to testing the LRH, and

Figure 10
Disfluency Rates for Speech With Spatial Content and With Nonspatial Content and When Subjects Were and Were Not Allowed to Gesture



Note. Reprinted from "Why do we gesture when we speak?", by R. M. Krauss, 1998. *Current Directions in Psychological Science*, *7*(2), p. 58. Copyright [1998] by SAGE Publications. Reprinted with permission.

it does not support the author's claim that "the effects of preventing gesturing depended on whether the conceptual content of the speech was spatial or nonspatial" (p. 229).[5]

Is it possible that Rauscher and colleagues (Rauscher et al., 1996) found the selective effect of gesture prevention on spatial speech, but simply failed to report the required two-way interaction? This was not the case; given the nonsignificance of the critical simple effect of preventing gesture in the spatial speech condition, the only way that the two-way interaction could become significant would be if preventing gesture had an unpredicted *facilitating* effect on speech production in the nonspatial speech condition. Thus, even if the nonreported (but critical) two-way interaction were significant, this interaction would not provide any clear support for the LRH since (a) it did not comprise the critical simple effect in the spatial speech condition, and (b) the statistical significance of the interaction would depend on a "backward" simple effect in the nonspatial speech condition, where no effect of gesture prevention was predicted (see the two bars on the right in Figure 10).

Speech Rate

Rauscher et al. (1996) reported analyses of speech rate as a second test of the LRH. As for disfluency rate, however, there was no clear evidence that the two required effects of gesture prevention on speech rate supported the LRH. For speech rate, the LRH predicted a critical simple effect showing that preventing gesture causes slower *spatial* speech. Looking at the results depicted in their plots, there is a trend consistent with this simple effect: Numerically, the speech rate was slower during speech with spatial content when participants were prevented from gesturing (striped bars), compared with when they were allowed to gesture (black bars, see the three pairs of bars on the left of Figure 8, reprinted from Rauscher et al., 1996). However, this trend may not be statistically significant; although the authors suggested that "with spatial content, speakers spoke more slowly when they could not gesture" (p. 228), no statistical test of this effect was reported, and no error bars were provided to guide interpretation of the trends.

Even if Rauscher et al. (1996) had obtained the critical simple effect, a further two-way interaction would be necessary to show that preventing gesture causes slower speech rates *selectively* during speech with spatial content. Rauscher et al. (1996) indeed reported a significant two-way interaction between gesture condition and speech content for speech rate. However, this interaction does not provide any clear support for the LRH since its composition appears problematic, in two ways: First, as noted above, there is no evidence that the required simple effect in the spatial speech condition was significant; second, the statistical significance of the interaction is driven in part by a trend toward a backward simple effect in the nonspatial speech condition, showing that people spoke *faster* when they were prevented from gesturing (i.e., see the three pairs of bars on the right of Figure 8 where the striped "no gesture" bars in Figure 8 are higher than the black "gesture" bars for speech with nonspatial content). Rauscher et al. (1996) acknowledged that a facilitating effect of preventing gesture on nonspatial speech was not predicted, writing: "when the content was nonspatial, speakers spoke more rapidly when they could not gesture. This latter result is puzzling to us, and we have no explanation for it" (p. 228).

In summary, the speech rate data from Rauscher et al. (1996) provide no clear support for the LRH. No statistical test was reported for the critical simple effect of gesture prevention on spatial speech. Although the critical two-way interaction was reported to be significant, (a) this interaction is necessary but not sufficient to support the LRH (absent the required simple effect in the spatial speech condition), and (b) its statistical significance is driven, in part, by the backward effect in the nonspatial speech condition.

Nonjuncture Filled Pause Rate

The third dependent measure reported by Rauscher et al. (1996) was the rate of nonjuncture filled pauses, which the authors suggest should be "the measure that most sensitively reflects problems in lexical retrieval" (p. 229) and is the most sensitive test of the LRH. On the basis of their analyses, Rauscher et al. (1996) claimed that "preventing gesturing increased the relative frequency of nonjuncture filled pauses in speech with spatial content, but not in speech with other content" (Rauscher et al., 1996, p. 226). This claim would require the same simple and interaction effects to be significant as in the previous dependent measures. In (partial) support of the LRH, the required simple effect was reported to be significant: Participants produced nonjuncture filled pauses at a higher rate when they were prevented from gesturing, compared with when they were allowed to gesture, during speech with spatial content. Yet, this simple effect is not sufficient to support the LRH. To support the claim that preventing gesture increases nonjuncture filled pauses *selectively*, "in speech with spatial content, but not in speech with other content," it would be necessary to present data in both the spatial and nonspatial speech conditions, and to test for the required 2-way interaction of gesture condition (gesture allowed; gesture prevented) and speech content (spatial content; nonspatial content). This interaction was not reported, and it cannot be tested in the reported data because Rauscher et al. (1996) did not report any results concerning nonjuncture filled pause rates in the nonspatial speech condition (Figure 9, reprinted from Rauscher et al., displays only spatial speech). As such, the authors did not present or analyze the required pattern of results needed to support their conclusion about the selective effect of preventing gesture on spatial speech.

Beyond Rauscher and Colleagues

Several studies in the 5 decades before and after Rauscher et al. (1996) have also failed to find reliable evidence for an increase in disfluent speech when speakers are prevented from gesturing (see Table 2). To our knowledge, no study has found higher disfluency rates, or higher nonjuncture filled pause rates, when speakers are prevented from gesturing than when they are allowed to gesture (Cravotta et al., 2018; Finlayson et al., 2003; Graham & Heywood, 1975; Hoetjes et al., 2014; Hostetter et al., 2007; Rimé et al., 1984). Only two studies (out of seven) report overall slower speech rates when individuals are prevented from gesturing than when they are

[5] This three-way interaction does not entail the critical two-way interaction (i.e., collapsing across the different speech conditions). Rather, the three-way interaction indicates that there was a difference across different speech conditions in whether or how the effects of gesturing depended on speech content. It could be the case, for example, that the effects of preventing gesturing depended on speech content only in the obscure and constrained conditions, and *not* in the natural speech condition that was critical for testing the LRH.

Table 2
Experiments That Tested Effects of Gesture Prevention on Disfluencies in Speech Production

Experiment	Graham and Heywood (1975)	Rimé et al. (1984)	Finlayson et al. (2003)	Morsella and Krauss (2004)	Hostetter et al. (2007)	Hoetjes et al. (2014)	Cravotta et al. (2018)
Participants	6 (all male)	16 (all male)	6 (all female)	79 (44 male, 35 female)	26 (19 female, 7 male)	38 (1/3 male)	10 (all female)
Speech task	Description of line drawings	Spontaneous conversation	Retelling a cartoon after watching a video	Description of visually present or absent objects	Description of three motor tasks (e.g., tying a shoe)	Description of videos showing how to tie a knot	Description of comic strips
Gesture manipulation	Keeping arms folded	Fastening arms to an armchair (also head, hands, legs, and feet)	Armchair fitted with strips of Velcro on the arms	Told electrodes on the arms are functional and would ruin the recording if moved	Wearing gloves adhered to a surface	Sitting on hands	Sitting on hands
Measures of disfluency in speech production [a]	(a) Number of hesitations (filled pauses and incomplete words), (b) hesitation rate (number of hesitations divided by total number of words), (c) number of pauses, (d) number of pauses excluding demonstratives, (e) speech rate (number of words divided by time spent speaking), (f) total time spent pausing, (g) proportion of total speaking time spent pausing excluding demonstratives, and (h) time spent pausing per line drawing	(a) Number of filled pauses, (b) number of incoherent sounds, and (c) speech rate (number of words per minute)	(a) Disfluency rate (number of unfilled and filled pauses, repetitions and reformulations per 100 words), (b) number of nonjuncture, and (c) juncture disfluencies	(a) Speech rate	(a) Number of filled pauses, (b) nonjuncture filled pauses rate (the proportion of filled pauses that were nonjuncture filled pauses), and (c) percentage of syntactic units starting with "and"	(a) Filled pause rate (number of filled pauses divided by number of words), and (b) speech rate	(a) Filled pause rate (number of filled pauses divided by the total number of words), (b) syntactical self-correction rate, (c) lexical self-corrections rate, (d) phonological self-correction rate, (e) repetition rate, (f) insertion rate, (g) interruption rate, (h) overall self-correction rate, (i) overall disfluency rate, (j) overall disfluency rate (including silent pauses), (k) overall disfluency rate (excluding filled and silent pauses), and (l) speech rate
Effect of preventing gesture on number of disfluencies or disfluency rate	Not significant for any of the number of disfluency measures (a, c, d) or for disfluency rate (b)	Not significant for any of the number of disfluency measures (a, b)	Not significant for disfluency rate (a)	N/A	Not significant for number of disfluencies (a)	Not significant for disfluency rate (a)	Not significant for any of the disfluency rate measures (a, b, c, d, e, f, g, h, I, j, k)
Effect of preventing gesture on speech rate	Not significant for speech rate (e)	Not significant for speech rate (c)	N/A	Significantly slower speech rate when prevented from gesturing compared with allowed to gesture (a)	N/A	Not significant for speech rate (b)	Significantly slower speech rate when prevented from gesturing compared with allowed to gesture (l)
Effect of preventing gesture on number of nonjuncture filled pauses	N/A	N/A	Not reported for number of nonjuncture disfluencies (b) Not significant for	N/A	Not significant for nonjuncture filled pause rate (b)	N/A	

(*table continues*)

Table 2 (*continued*)

Experiment	Graham and Heywood (1975)	Rimé et al. (1984)	Finlayson et al. (2003)	Morsella and Krauss (2004)	Hostetter et al. (2007)	Hoetjes et al. (2014)	Cravotta et al. (2018)
or nonjuncture filled pause rate			number of juncture disfluencies (c)				
Effect of preventing gesture on other measures of disfluency	Not significant for total time spent pausing (f), or time spent pausing per line drawing (h) There was a significant effect of preventing gesture only on the proportion of total speaking time spent pausing excluding demonstratives (g)	N/A	N/A	N/A	Significantly higher percentage of units starting with "and" when prevented from gesturing compared with allowed to gesture (c)	N/A	N/A

[a] Disfluency measures are listed for which a separate statistical test is conducted such that number of disfluency measures in this list corresponds to the number of comparisons in the experiment.

allowed to gesture (Cravotta et al., 2018; Morsella & Krauss, 2004). However, in one of these studies (Cravotta et al., 2018), the finding that speech was slower when gesture was prevented was the *only* significant comparison among a total of 22 comparisons, which included other disfluency measures; therefore, this single significant finding would not remain statistically significant after correcting the alpha-level for multiple comparisons. Morsella and Krauss (2004) also found slower speech rates when gesture was prevented than when it was allowed. However, the authors did not find a selective effect of gesture prevention on speech rate in one experimental condition more than in another; thus, it is unclear whether the results simply reflect a particularly restrictive gesture-prevention method (i.e., participants had electrodes placed on their forearms and were instructed not to move because "movement of the limbs could ruin the quality of the recordings"; p. 89). Notably, none of the studies reporting null effects of gesture prevention distinguished between disfluencies during spatial and nonspatial speech; thus, arguably, these studies did not attempt to validate Rauscher et al.'s (1996) claim that gesture prevention selectively affects spatial speech. By contrast, our study tested this claim explicitly, but still found no evidence that gesture prevention increases disfluency—in either spatial or nonspatial speech.

Beyond Disfluency: Is There Any Evidence That Preventing Gestures Hurts Speech Production?

Given that there is no reliable evidence that gestures help people produce fluent speech, is there *any* evidence that gestures facilitate speech production in other ways? In principle, preventing gesture could hurt some aspect of speech production that does not result in speech disfluencies. For example, according to Rimé et al. (1984), preventing people from gesturing lowered the vividness of the imagery in speech. However, a reexamination of the analyses reported in Rimé et al. (1984) indicates that this reported effect should not be interpreted as statistically significant production (see column 3 on Table 2 for the dependent variables tested in Rimé et al., 1984). The results showed a p-value less than .05 for only one these six variables (effect of gesture prevention on imagery index: $F(4, 48) = 3.19$, $p = .02$). This result is widely cited as evidence that gestures facilitate speech production by activating semantic features that enter into word search (e.g., see Rauscher et al., 1996, p. 226). Yet, interpreting this one result in the context of Rimé et al.'s full study would require correcting the alpha level for multiple comparisons. A Bonferroni-corrected alpha-value for six independent tests would be $\alpha = .008$; Rimé et al.'s (1984) reported effect of gesture prevention on imagery would not approach significance based on this corrected alpha-value.

Another set of studies testing effects of gesture prevention on speech, beyond disfluencies, examined whether people find it harder to generate a target word from definitions or from pictures when prevented from gesturing. This effect was tested in three papers (Beattie & Coughlan, 1999; Frick-Horbury & Guttentag, 1998; Pine et al., 2007). In influential papers reviewing these findings (e.g., see Cook et al., 2010; Hostetter & Alibali, 2008; Kita et al., 2017; Wesp et al., 2001), two out of the three studies have been cited as evidence that gestures help people find the right words; only one study, by Beattie and Coughlan (1999), has been cited as showing no evidence in support of this view.[6] Our reexamination of these three studies revealed that, in fact, none of these papers shows consistent support for the hypothesis that gesturing helps people to find target words. Below, we examine the pattern of results that led Beattie and Coughlan (1999) to conclude that they found no evidence in support of the LRH, and we show that according to the same criteria, the two other studies showed no clear evidence in support of the LRH (Frick-Horbury & Guttentag, 1998; Pine et al., 2007; see Table 3 for a detailed comparison of the three studies). On the contrary, Frick-Horbury and Guttentag's (1998) study offered the opposite conclusion: "In the present study, however, there was little evidence that gesture production per se enhanced verbal recall" (p. 54).

What were the patterns of results that led Beattie and Coughlan (1999) to conclude that gestures did not help people find target words, as had been predicted? Beattie and Coughlan (1999) found no statistically significant evidence that people remember fewer words when prevented from gesturing, compared with when they were allowed to gesture—this null effect in Beattie and Coughlan (1999) has been acknowledged by subsequent papers summarizing these results (see sixth row on Table 3). However, this null effect of gesture prevention on word finding was only one among a

[6] For example, Kita et al. (2017) summarized the findings of these three studies by stating that "Frick-Horbury and Guttentag (1998) and Pine et al. (2007) reported evidence that gesture facilitates lexical retrieval, but Beattie and Coughlan (1999) did not."

Table 3
Experiments That Tested the Relationship Between Gestures and Word Finding, Cued by Definitions or Pictures

Experiment	Frick-Horbury and Guttentag (1998)[a]	Beattie and Coughlan (1999)	Pine et al. (2007)
Participants	36 (11 male and 25 female)	60 (20 male and 40 female)	65 children (33 boys and 32 girls)
Speech task	Retrieval: Finding the target words from definitions for a set of 50 low frequency words. Recall: Listing as many words as one can remember from the list of words on the retrieval task.	Finding the target words from definitions for a set of 25 low frequency words.	Finding the target words from pictures for two sets of 25 pictures to name.
Gesture manipulation	Holding a rod with both hands (between-subject)	Keeping them folded so as to prevent any gesturing (between-subject)	Hands placed in mittens that stuck to a board in front of them (within-subject)
Definition of a tip-of-the-tongue (TOT) state	Participants are asked to report when they experience the phenomenon defined as "knowing you know a word that you are unable to generate."	A trial was coded as one with a TOT state when the participant displayed behaviors such as saying "Oh, God I know it!," wincing, etc.	A trial was coded as one with a TOT state when the participant displayed behaviors such as saying "I knew the word but I couldn't take it out," wincing, etc.
Does gesture prevention make people remember fewer target words?	*Yes* *Significantly fewer target words were retrieved accurately when prevented from gesturing compared with allowed to gesture.* *Significantly fewer target words were recalled accurately when prevented from gesturing compared with allowed to gesture.*	No statistically significant evidence. Numerically a higher proportion of target words were retrieved accurately when prevented from gesturing compared with allowed to gesture. Numerically a higher proportion of target words were retrieved accurately without a cue when prevented from gesturing compared with allowed to gesture.	*Yes* *Significantly fewer pictures were named accurately with the target word when prevented from gesturing compared with allowed to gesture.*
Is gesturing associated with remembering the target word?	**No**[b] **Numerically less likely to remember the target word when there is gesture (21% of words remembered) compared with when there is no gesture (53% of words remembered)**. Numerically less likely to remember the target word when there is gesture (19% of words remembered) compared with when there is no gesture (23% of words remembered), when trials that are too short are excluded.	N/A	*Yes*[c, e] *Significantly more gestures were produced before remembering the target word compared with before failing to remember the target word.*
Is iconic gesturing *selectively* associated with remembering the target word?	N/A	N/A	**No**[e] **Significantly more iconic gestures, but also significantly more beat gestures and self-adaptors, were produced before remembering the target word compared with before failing to remember the target word.**
Does gesture prevention make it harder to resolve TOT states?	N/A[d]	*Yes* *Significantly lower rate of TOT states was resolved when prevented from gesturing compared with allowed to gesture.*	*Yes* *Significantly lower rate of TOT states was resolved when prevented from gesturing compared with allowed to gesture.*
Is gesturing associated with resolving TOT states?	N/A[d]	**No** **Significantly lower rate of TOT states was resolved when there is gesture compared with when there is no gesture.**	*Yes*[e] *Significantly higher number of gestures were produced when TOT states were resolved compared with when TOT states were not resolved.*
Is iconic gesturing *selectively* associated with resolving TOT states?	N/A[d]	No statistically significant evidence. Numerically lower rate of TOT states was resolved when there is iconic gesture compared with when there is no iconic gesture—same was true for beat gestures and self-adaptors.	**No**[e] **Significantly higher number of iconic gestures, but also beat gestures and self-adaptors, were produced when TOT states were resolved compared with when TOT states were not resolved.**

(*table continues*)

Table 3 (*continued*)

Experiment	Frick-Horbury and Guttentag (1998)[a]	Beattie and Coughlan (1999)	Pine et al. (2007)
Does gesture prevention cause people to experience more TOT states?	No statistically significant evidence. Numerically similar number of TOT states were experienced when prevented from gesturing compared with allowed to gesture.[f]	**No.** **Significantly fewer TOT states were experienced when prevented from gesturing compared with allowed to gesture.**	No statistically significant evidence. Numerically fewer number of TOT states were experienced when prevented from gesturing compared with allowed to gesture.
Is gesturing associated with experiencing TOT states?	*Yes*[b] *Numerically more likely to gesture during trials with TOT states (82%) compared with trials without TOT states (non-TOT missed trials: 9%; non-TOT correct trials: 16%).*	N/A	*Yes*[e] *Significantly higher number of gestures were produced during trials with TOT states compared with trials without TOT states.*
Is iconic gesturing *selectively* associated with experiencing TOT states?[g]	*Yes* *Significantly higher number of iconic gestures were produced than vague gestures or motor movements during trials with TOT states.*	No statistically significant evidence. Numerically fewer iconic gestures were produced than both beat gestures and self-adaptors during trials with TOT states.	*Yes*[e] *Significantly higher number of iconic gestures were produced during trials with TOT states compared with trials without TOT states, but this was not true for beat gestures and self-adaptors.*

Note. Descriptions in italics indicate a statistically significant effect in support of the LRH. Descriptions in regular indicate null results. Descriptions in bold indicate statistically significant effects contradicting the Lexical Retrieval Hypothesis (LRH).
[a] Frick-Horbury and Guttentag (1998) conducted two experiments. However, here we report results from the first experiment only, since the results and analyses of the second experiment are not reported. [b] Frick-Horbury and Guttentag (1998) did not report statistical tests for comparing proportions across experimental conditions and made conclusions based on numerical results. Therefore, we report here based on their interpretation of the significance of the differences across experimental conditions and provide the numerical results across experimental conditions. [c] Pine et al. (2007) did not report a statistical test for this result; however, given that all of the subtypes of gestures showed a statistically significant difference across the experimental conditions, this result for gestures overall must be by necessity statistically significant. [d] Frick-Horbury and Guttentag (1998) aimed to test for differences in TOT resolution rate across experimental conditions. However, they obtained a floor effect for the overall rate of resolved TOT states (only 2% of all TOT states were resolved). Therefore, no meaningful comparisons across experimental conditions were possible. [e] Note the use of different measures to address this conceptual question across the three studies. [f] Frick-Horbury and Guttentag (1998) did not report the summary statistics for this measure, so we cannot report on the numerical comparison. [g] Note the use of different measures to address this conceptual question across the three studies.

broader set of results that led Beattie and Coughlan (1999) to conclude that their study showed no evidence in support of the LRH (see third column on Table 3), and it was not even the effect that the authors considered to be the most critical (see p. 41 and 50). They also tested whether gesture prevention makes it harder for people to resolve tip-of-the-tongue states (TOT states): situations in which speakers know the gist of what they want to say, but they cannot find the right word. Beattie and Coughlan (1999) *did* find statistically significant evidence that gesture prevention makes it harder to resolve TOT states, but they did not interpret this result as supporting the LRH. Critically, they argued that an effect of gesture prevention on resolving TOT states is not clear evidence for the LRH. Why not? Because only *iconic* gestures are hypothesized to "have a functional role in lexical access" (Beattie & Coughlan, 1999, p. 35; see also Butterworth & Hadar, 1989). If an analysis includes all gesture types, then the effects of gesture prevention on resolving TOT states "may, of course, have nothing to do with the occurrence of *iconic* gestures" (p. 41, italics added). In fact, there was no evidence to suggest that Beattie and Coughlan's (1999) effect was driven by the prevention of *iconic* gestures; therefore, the effect did not support the LRH (or related proposals, e.g., Butterworth & Hadar, 1989).

Beyond the ambiguity of their gesture *prevention* results, Beattie and Coughlan (1999) tested for two critical links between gesture *production* and word finding, both of which failed to support the LRH. First, if gestures help people resolve TOT states, then people should resolve more TOT states on trials where they produced a gesture, compared with trials where they did not produce a gesture. However, Beattie and Coughlan (1999) found the opposite pattern: Only 64% of the TOT states were resolved when people produced a gesture, whereas *all* of the TOT states were resolved when people did *not* produce a gesture. Second, the LRH predicts that only *iconic* gestures should help people resolve TOT states. If this is true, then more TOT states should be resolved when people produced iconic gestures, compared with when they did not produce any iconic gestures. Yet, there was no statistically significant evidence that people resolved more TOT states when they produced an iconic gesture (69%), compared with when they did not (73%); rather, there was a numerical trend in the opposite direction. Overall, even though Beattie and Coughlan (1999) obtained *one* result predicted by the LRH, a main effect of gesture prevention on TOT resolution, they concluded that "if one considers the broader context here, there is no real evidence that gesturing facilities lexical access" (p. 47, see third column Table 3 for only one result among nine results that supports the LRH).

Although Frick-Horbury and Guttentag (1998) is widely cited as evidence for the LRH, the authors stated that their study does not provide clear support for the LRH because they found that the occurrence of gestures was not actually associated with successful word finding. Frick-Horbury and Guttentag (1998) *did* find a main effect of gesture prevention: People remembered fewer words when they were prevented from gesturing, compared with when they were allowed to gesture. However, like Beattie and Coughlan (1999), Frick-Horbury and Guttentag (1998) argued that this effect of gesture prevention on word finding is not clear evidence for the LRH. To determine whether their results supported the LRH, they tested whether gesture production was associated with successful word finding. Like Beattie and Coughlan (1999), Frick-Horbury and Guttentag (1998) found the opposite pattern: Only 21% of the words were correctly remembered when people produced a gesture, whereas 53% of the words were correctly remembered when people did *not* produce a gesture. Based on these results, Frick-Horbury and Guttentag (1998) concluded that "It is possible that a focus on overt gesture production is the wrong place to look for an explanation for the hand restriction effect."[7]

Using the same criteria that Beattie and Coughlan (1999) and Frick-Horbury and Guttentag (1998) used in evaluating their results, we can conclude that Pine et al. (2007) also did not provide clear evidence in support of the LRH. Pine et al. (2007) *did* find two main effects of gesture prevention: Children remembered fewer words, and resolved fewer TOT states, when they were prevented from gesturing, compared with when they were allowed to gesture. And unlike the other studies, Pine et al. (2007) *did* also find that gesture production may be responsible for these effects: Children gestured more before successfully finding a target word and before successfully resolving a TOT state. However, critically, they showed no evidence to suggest that these effects were selectively driven by the production of iconic gestures: Children used more iconic gestures, but also more beat gestures and more self-adaptors (i.e., simple self-touching movements), when they were able to find a word (i.e., remembering the target word, resolving a TOT state), compared with when they failed to find a word (i.e., not remembering the target word, failing to resolve a TOT state). Given that the association of more gestures with successful word finding was not selective to iconic gestures, iconic gesture production may be the wrong place to look for an explanation of the gesture prevention effects. Being allowed to make *any* movement, including self-adaptors, may make children feel more comfortable, which could help them succeed in the experimental task. Therefore, Pine et al.'s (2007) results also do not provide clear evidence for the LRH.

So far, we have focused our reexamination of Beattie and Coughlan (1999), Frick-Horbury and Guttentag (1998), and Pine et al. (2007) on only two measures of word finding: the number of target words remembered, and the number of TOT states resolved. However, all three studies also tested the effect of gesture prevention on a third measure of word finding: whether people *experience* more TOT states when they are prevented from gesturing, compared with when they are allowed to gesture. If gestures help people find the right words, people should experience more TOT states when gestures are not available to them. However, none of the three studies showed evidence supporting this prediction (see 12th row on Table 3). Furthermore, Beattie and Coughlan (1999) found statistically significant evidence in the opposite direction of this prediction: People experienced *fewer* TOT states when they were prevented from gesturing, compared with when they are allowed to gesture. Overall, our reexamination revealed that none of these studies showed clear support for the LRH; on the contrary, some measures showed consistent evidence against the LRH (see rows 11 and 12 of Table 3).

To conclude, our reexamination of the studies summarized in Tables 2 and 3 corroborates our reexamination of Rauscher et al. (1996): There appears to be no clear evidence that preventing gesture hurts speech production—by making speech more disfluent, by lowering the imagery level in speech, by hindering word finding and TOT state resolution, or by making people experience more TOT states. Overall, we conclude that there is no clear support for the long-standing, influential claim that people gesture when they speak, in part, because gesturing helps speakers produce the right words.[8]

Why Do People Gesture When They Speak?

Krauss (1998) offered the LRH as an answer to the question, "Why do we gesture when we speak?" A "why" question explores the function of a behavior: What consequences follow a behavior that leads to that behavior's recurrence? The LRH provided the first functional explanation for what causes a speaker to gesture that did not restrict cospeech gesture's function to communicative benefits for the listener, and instead suggested that gesturing may serve a cognitive function in the speaker's mind.

Even though gesturing does not facilitate speech production, gestures do appear to serve a variety of other speaker-internal cognitive functions. For example, gesturing lightens speakers' cognitive load and frees up cognitive resources that can be allocated to other tasks: Gesturing while explaining a math task allowed speakers to remember more words on a simultaneously performed word recall task (Goldin-Meadow et al., 2001; Ping & Goldin-Meadow, 2010; Wagner et al., 2004). Gesturing also facilitates learning: Gesturing while learning a new math concept helps learners generalize (Goldin-Meadow et al., 2009; Novack et al., 2014; Wakefield et al., 2018) and retain (Cook et al., 2008) the knowledge they gained during instruction.

As Krauss (1998) argued, gesture's function cannot be limited to communicative purposes. A complete functional account of why people gesture needs to include not only communicative functions for the listener, but also speaker-internal cognitive functions. In light of the evidence we present here, such an account of the cognitive functions of gesture will need to look beyond the role of gesture in speech production.

Conclusions

Do gestures facilitate speech production? Typically, reviews listing the cognitive functions of gesture start this list with the assertion that gesture helps people find the right words (e.g., Alibali, 2005; Goldin-Meadow, 1999; Kita et al., 2017). Here, however, we showed that there is no compelling evidence to support this influential hypothesis. Rauscher et al.'s (1996) claim that gesture prevention makes spatial speech disfluent is among the most widely cited empirical results in the gesture literature; yet, upon reexamining this study, we found that the data do not support the LRH. Our further reexamination of 5 decades of research testing versions of this hypothesis, before and after Rauscher et al.'s (1996) influential study, revealed that there is no reliable evidence that gesture prevention hurts speech production. Accordingly, the results from our study showed no statistically significant effects of gesture prevention on speech disfluency, for speech about literal or

[7] Frick-Horbury and Guttentag (1998) aimed to also test the effect of gesture prevention on another measure of word finding, that is TOT resolution; however, the overall rate of TOT states resolved was very low (2%) that prevented meaningful comparison across experimental conditions.

[8] We thank a reviewer for raising the possibility that LRH might explain why people gesture when they speak only for people with fragile lexical retrieval (e.g., children or adults learning a second language). If so, the LRH could be helpful for understanding speech production under special conditions, but would not provide an answer to the question we sought to address in the current manuscript: Why do people gesture when they speak in general (even with no compromised lexical retrieval)?

metaphorical space. Gestures do *not* appear to facilitate speech production, challenging long-held beliefs about why people gesture when they speak.

References

Agresti, A. (2003). *Categorical data analysis* (Vol. 482). Wiley.

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, *5*(4), 307–331. https://doi.org/10.1207/s15427633scc0504_2

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, *15*(6), 593–613. https://doi.org/10.1080/016909600750040571

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138–1144. https://doi.org/10.1177/0956797611417722

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology*, *90*(1), 35–56. https://doi.org/10.1348/000712699161251

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, *96*(1), 168–174. https://doi.org/10.1037/0033-295X.96.1.168

Casasanto, D. (2008). Similarity and proximity: When does close in space mean close in mind? *Memory & Cognition*, *36*(6), 1047–1056. https://doi.org/10.3758/MC.36.6.1047

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367. https://doi.org/10.1037/a0015854

Casasanto, D. (2013). Gesture and language processing. In H. Pashler, T. Crane, M. Kinsbourne, F. Ferreira, & R. Zemel (Eds.), *Encyclopedia of the mind* (pp. 372–374). Sage. https://doi.org/10.4135/9781452257044.n138

Casasanto, D., & Bottini, R. (2014). Spatial language and abstract concepts. *Wiley Interdisciplinary Reviews. Cognitive Science*, *5*(2), 139–149. https://doi.org/10.1002/wcs.1271

Cienki, A. (2005). Image schemas and gesture. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (Vol. 29, pp. 421–442). Mouton de Gruyter. https://doi.org/10.1515/9783110197532.5.421

Clark, H. H. (1973a). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Clark, H. H. (1973b). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and acquisition of language* (pp. 27–63). Academic Press. https://doi.org/10.1016/B978-0-12-505850-6.50008-6

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047–1058. https://doi.org/10.1016/j.cognition.2007.04.010

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language*, *63*(4), 465–475.

Cravotta, A., Busà, M. G., & Prieto, P. (2018). Restraining and encouraging the use of hand gestures: Effects on speech. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. S¨ledziński (Eds.), *Proceedings 9th International Conference on Speech Prosody* (pp. 206–210). Poznań, Poland.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Finlayson, S., Forrest, V., Lickley, R., & Beck, J. M. (2003). Effects of the restriction of hand gestures on disfluency. *Proceedings of Diss, Gothenburg Papers in Theoretical Linguistics*, *90*, 21–24.

Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American Journal of Psychology*, *111*(1), 43. https://doi.org/10.2307/1423536

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

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–272. https://doi.org/10.1111/j.1467-9280.2009.02297.x

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, *5*(2), 189–195. https://doi.org/10.1002/ejsp.2420050204

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. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). Does sitting on your hands make you bite your tongue? The effects of gesture prohibition on speech during motor descriptions. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the Cognitive Science Society* (pp. 1097–1102). Erlbaum.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, *396*(6708), 228. https://doi.org/10.1038/24300

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*(2), 54–54. https://doi.org/10.1111/1467-8721.ep13175642

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell & L. Messing (Eds.), *Gesture, speech, and sign* (pp. 93–116). Oxford University Press. http://www.columbia.edu/~rmk7/PDF/K&H.pdf.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208. https://doi.org/10.1207/s15516709cog0402_4

Matthews, J. L., & Matlock, T. (2011). Understanding the Link Between Spatial Distance and Social Distance. *Social Psychology*, *42*(3), 185–192. https://doi.org/10.1027/1864-9335/a000062

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

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, *117*, 411–424. https://doi.org/10.2307/4149008

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software*, *43*(9), 1–14. https://doi.org/10.18637/jss.v043.i09

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, *25*(4), 903–910. https://doi.org/10.1177/0956797613518351

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747–754. https://doi.org/10.1111/j.1467-7687.2007.00610.x

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, *34*(4), 602–619. https://doi.org/10.1111/j.1551-6709.2010.01102.x

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

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*(4), 311–325. https://doi.org/10.1007/BF00991870

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*(2), 113–129. https://doi.org/10.1016/j.cognition.2005.01.004

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*(1), 1–21. https://doi.org/10.1037/0022-3514.89.1.1

Wagner, S., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*, 395–407. https://doi.org/10.1016/j.jml.2004.01.002

Wakefield, E. M., Hall, C., James, K. H., & Goldin-Meadow, S. (2018). Gesture for generalization: Gesture facilitates flexible learning of words for actions on objects. *Developmental Science*, *21*(5), e12656. https://doi.org/10.1111/desc.12656

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, *114*(4), 591.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of the 5th International Conference on Language Resources and Evaluation* (pp. 1556–1559). European Language Resources Association.

Yap, D., Brookshire, G., & Casasanto, D. (2018, July). *Beat gestures encode spatial semantics* [Paper presentation]. 40th Annual Conference of the Cognitive Science Society, Madison, WI.

Received December 22, 2020
Revision received July 8, 2021
Accepted July 28, 2021 ■