

Gesture helps, only if you need it:

Inhibiting gesture reduces tip-of-the-tongue resolution for those with weak short-term memory

Jennie E. Pyers¹, Rachel Magid^{1,2}, Tamar H. Gollan³, Karen Emmorey⁴

¹Wellesley College, Psychology Department, 106 Central St., Wellesley, MA 02481, jpyers@wellesley.edu [Corresponding Author]

²Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, rwmagid@mit.edu

³Department of Psychiatry, University of California, San Diego, 9500 Gilman Drive #0948 La Jolla, CA 92093 tgollan@ucsd.edu

⁴Laboratory for Language and Cognitive Neuroscience, San Diego State University, 6495 Alvarado Road, Suite 200, San Diego, CA 92120, kemmorey@sdsu.edu

Keywords: Lexical retrieval; tip-of-the-tongue states, gesture; verbal short-term memory; spatial short-term memory; individual differences

Abstract

People frequently gesture when a word is on the tip of their tongue (TOT), yet research is mixed as to whether and why gesture aids lexical retrieval. We tested three accounts: the lexical retrieval hypothesis, which predicts that semantically related gestures facilitate successful lexical retrieval; the cognitive load account, which predicts that matching gestures facilitate lexical retrieval only when retrieval is hard, as in the case of a TOT; and the motor movement account, which predicts that any motor movements should support lexical retrieval. In Experiment 1 (a between-subjects study; $N=90$), gesture inhibition, but not neck inhibition, affected TOT resolution but not overall lexical retrieval; participants in the gesture inhibited condition resolved fewer TOTs than participants who were allowed to gesture. When participants could gesture, they produced more representational gestures during resolved than unresolved TOTs, a pattern not observed for meaningless motor movements (e.g., beats). However, the effect of gesture inhibition on TOT resolution was not uniform; some participants resolved many TOTs, while others struggled. In Experiment 2 (a within-subjects study; $N=34$), the effect of gesture inhibition was traced to individual differences in verbal, not spatial short-term memory (STM) span; those with weaker verbal STM resolved fewer TOTs when unable to gesture. This relationship between verbal STM and TOT resolution was not observed when participants were allowed to gesture. Taken together these results fit the cognitive load account; when lexical retrieval is hard, gesture effectively reduces the cognitive load of TOT resolution for those who find the task especially taxing.

Keywords: Lexical retrieval; tip-of-the-tongue states, gesture; verbal short-term memory; spatial short-term memory; individual differences

Gesture helps, only if you need it:**The relationship between gesture, short-term memory, and lexical retrieval****1. Introduction**

Gesture is ubiquitous during speech and seems to play a significant role in the perception and production of language (e.g., McNeil, 1985); however, whether gesture plays a specific role in lexical retrieval is still unclear. Although speakers anecdotally report gesturing more when they have difficulty retrieving a word that is on the tip of their tongue (TOT), their gestures may do little to help them retrieve the word. For example, TOT states may elicit more gestures because in conversation, speakers use gesture as a signal that they are not yet ready to yield their conversational turn (Duncan, 1972), and therefore they may gesture as a strategy to keep their listeners from interrupting while they search for the right word. Indeed, speakers gesture more during a lexical retrieval task with an interlocutor present than when sitting alone speaking into a recorder (Holler, Turner, & Varcianna, 2012). On the other hand, because TOTs reflect inadequate activation of the target word and often involve incorrect selection of a non-target candidate (Brown, 1991), speakers' gestures might boost the activation of the correct lexical representation, making the target word more available for retrieval (e.g., Frick-Horbury & Guttentag, 1998). Co-speech gesture may also provide a strategy for people to offload the cognitive work of lexical search (e.g., Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow, 2003). The goal of the current study was to identify whether and under what conditions gesture plays a role in lexical retrieval or in helping to resolve TOT states.

Three accounts have been proposed to explain the relationship between gesture and lexical retrieval. The first account—the lexical retrieval hypothesis—posits that the role of gesture in speech production is specific to lexical retrieval and plays little role in other aspects of language production or cognitive processing (Rauscher, Krauss, & Chen, 1996). According to the lexical retrieval hypothesis, gesture cross-modally primes the phonological elements of a word via spatial and semantic pathways, making it more accessible for retrieval (Rauscher et al., 1996). In this case, prior to the preverbal message, spatial/dynamic information that overlaps with the propositional information to be conveyed is transmitted from working memory through the motor system in parallel with lexical selection. Once a “lexical” movement is generated (i.e., a gesture affiliated with a word), it can prime either the selection of the word concept, the lemma, which captures *what* you want to name, or the selection of the word form, which are the

phonological features of the word. Under either case, “lexical” movements facilitate the production of the target word (Krauss & Hadar, 1999). According to this argument, representational gestures, which encode semantic features of a word (sometimes called “lexical gestures,” e.g., Krauss & Hadar, 1999), support the retrieval of spatial words while non-iconic gestures, such as beats, do not (see McNeill, 1985, for a review of gesture types). Although Krauss and Hadar (1999) specifically suggest that it is meaningless to break down representational gestures into more fine-grained categories such as iconic gestures, which represent a referent, and metaphoric gestures, which represent elements of abstract ideas (see McNeill, 1995), the mechanism that they argue underlies gestural priming of a word rests on explicit form-referent similarity encoded in the gesture. That is, the relevant semantic feature of the gesture is what ultimately primes the target word. As such, iconic gestures with a high degree of semantic overlap with the target word should better support lexical retrieval.

The cognitive load account adopts a broader perspective suggesting that gesture helps with any difficult task, including but not exclusively, lexical retrieval, by lifting the cognitive burden (Goldin-Meadow, 2003). Gesture can support lexical retrieval, verbal message planning (e.g., Information Packaging Hypotheses; Kita, 2000), and acquiring and processing difficult concepts (e.g., Gunderson, Spaepen, Gibson, Goldin-Meadow, & Levine, 2015; Trofatter, Kontra, Beilock, & Goldin-Meadow, 2014). Here the hypothesis is that when we wrestle with a difficult problem, our gestures can lighten the load on working memory—the ability to store information while engaging in an additional processing task (Cook et al., 2012). While the precise mechanism of how gesture lightens the load is unclear (see Goldin-Meadow, 2003, for a review of possibilities), some researchers propose that rather than priming the target word, representational gestures help maintain visual imagery in working memory during lexical search (e.g., Wesp, Hesse, Keutmann, & William, 2001). Like the lexical retrieval hypothesis, the cognitive load account suggests that “matching gestures” which match the content of speech can support processing and recall within both linguistic and nonlinguistic domains (Wagner, Nusbaum, & Goldin-Meadow, 2004). Matching gestures are not necessarily representational; gestures that do not appear representational (e.g., deictic pointing gestures) nonetheless can convey information that is complementary to the speech (e.g., Wagner, et al, 2004). Therefore, within this framework, gestures that serve to lighten the cognitive load can be iconic, metaphoric, or deictic. As is the case under the lexical retrieval hypothesis, the cognitive load

account explicitly excludes non-meaningful movements, such as beats (e.g., rhythmic gestures that carry no semantic content) or self-adaptors (grooming-type gestures), as playing a role in facilitating lexical or cognitive processing (e.g. Cook et al, 2012).

One key difference between the lexical retrieval hypothesis and the cognitive load account may be the locus of the effect of gesture on lexical retrieval. The lexical retrieval hypothesis suggests that the effect of gesture is at either lemma or word form selection. In contrast, the cognitive load account suggests that gesture primarily plays a role when a task is difficult. In the case of lexical retrieval we can see this difficulty during TOTs, which arise after lemma selection but before the correct selection of the word form. Thus, the cognitive load account predicts that the greatest effect of gesture on lexical retrieval will not be on the number of correct retrievals but on the eventual resolution of a TOT. In contrast, the lexical retrieval hypothesis predicts that gesture will aid both in correct retrievals as well as in the resolution of TOTs.

Similar evidence could be used to support both the lexical retrieval hypothesis and the cognitive load account in the case of lexical retrieval. Inhibition of gesture affects the fluency of language with spatial content in particular (Rauscher et al., 1996), and it specifically affects lexical retrieval, as evidenced by the finding that participants who are prevented from gesturing retrieve fewer words (Frick-Horbury & Guttentag, 1998) and resolve fewer TOTs compared to when they are allowed to gesture (Beattie & Coughlan, 1999; Pine, Bird, & Kirk, 2007, although see Frick-Horbury & Guttentag, 1998 for a null finding). Children also produce more iconic gestures, but not beats or self-adaptors, during TOT states compared to non-TOT states (Pine et al., 2007).

Yet, several findings speak against both accounts. Beattie and Coughlan (1999) found that speakers prevented from gesturing experienced a similar number of successful retrievals compared to speakers' who were allowed to gesture, and when allowed to gesture, speakers produced fewer iconic gestures than other types of gestures during TOT states. Thus, increased production of iconic gestures did not improve their TOT resolution rates (see also Frick-Horbury & Guttentag, 1998). In fact, some studies show that speakers have better lexical retrieval when they do not spontaneously produce a gesture compared to when they do (Beattie & Coughlan, 1999; Frick-Horbury & Guttentag, 1998). However, this finding could be attributed to the fact that successful retrieval occurs within a small time frame often before speakers have the

opportunity to produce a gesture. Another finding that does not support these accounts involves the semantic fluency task, which measures a speaker's efficiency in accessing the lexicon (participants name as many items in a semantic category, e.g. food, as they can in a limited time frame, such as 60 seconds). Studies have found that performance on a semantic fluency task does *not* correlate with a speaker's propensity to use representational gestures (Gillespie, James, Federmeier, & Watson, 2014; Hostetter & Alibali, 2007), indicating that gesture production may not be linked to efficient lexical access.

While representational gestures may not specifically support successful TOT resolution (e.g., Beattie & Coughlan, 1999), at least one study has shown that hand movements alone (self-paced tapping) seem to support TOT resolution for both spatial and non-spatial lexical items when participants are asked to type their responses (Experiments 1 & 2, Ravizza, 2003). Thus, a third account that considers gesture's role in lexical retrieval suggests that motor movement in general can aid lexical retrieval once lemma selection has taken place, but before the word-form is accessed. Movements of the hand may help by activating neural areas common to both language and motor production (Ravizza, 2003). Further evidence for this account comes from Hostetter and Alibali (2007) who observed a positive correlation between efficient lexical access, as measured by semantic fluency, and the use of beat gestures during a narration task. This finding supports the idea that the production of non-meaningful hand gestures could aid lexical retrieval.

When looking at the literature as a whole, what is clear is that the role of gesture in lexical retrieval is definitively unclear. The contradictory findings across studies could be attributed to methodological variability in eliciting TOTs, e.g., narrative descriptions (Rauscher et al., 1996), picture naming (Pine et al., 2007), or definitions (Beattie & Coughlin, 1998; Frick-Horbury & Guttentag, 1998; Holler et al., 2012; Ravizza, 2003); in defining a TOT, e.g., self-report (Frick-Horbury & Guttentag, 1998; Ravizza, 2003) or experimenter judgment (Beattie & Coughlin, 1999; Holler et al., 2012; Pine et al., 2007); or in identifying self-resolved TOTs, e.g., any resolution after any hesitation (Beattie & Coughlin, 1999) or only resolutions that occur after a specified window of time (Ravizza, 2003). Further, no between-group study that experimentally manipulated participants' ability to gesture has matched participants on variables such as age and education (e.g., Beattie & Coughlin, 1999; Frick-Horbury & Guttentag, 1998; Morsella & Krauss, 2004), which are known to affect lexical retrieval success (Gollan & Brown,

2006). In addition, studies varied in how they elicited and operationalized different gesture types, and which gesture types they included in their analyses. Some examined the effect of gestures produced during a lexical retrieval task to performance on that task (e.g., Beattie & Coughlin, 1999; Frick-Horbury & Guttentag, 1998; Pine et al., 2007), while others compared gesture use during a narrative task to performance on a different measure of lexical retrieval (Gillespie et al., 2015; Hostetter & Alibali, 2007). Some researchers defined representational/iconic gestures in the broadest sense of the term, combining transparently semantically-related gestures with metaphoric gestures (Rauscher et al, 1996), while others narrowly defined representational gestures as depicting only closely related semantic information (Beattie & Coughlin, 1999; Frick-Horbury & Guttentag, 1998; Gillespie et al., 2015; Holler et al., 2012; Smithson & Nicoladis, 2013). Some, but not all, studies included separate analyses for beats and self-adaptors (Beattie & Coughlin, 1999; Frick-Horbury & Guttentag, 1998; Hostetter & Alibali, 2007; Pine et al., 2007), while others included analyses for “word search gestures” (Holler et al, 2012). Only one study included deictic gestures in their set of “meaningful gestures” (Hostetter & Alibali, 2007), despite their role in supporting other aspects of language and cognitive processing (e.g., Wagner, et al, 2004). Finally, studies varied in whether they analyzed the raw number of gestures (Beattie & Coughlin, 1999; Frick-Horbury & Guttentag, 1998; Pine et al., 2007) or gesture rates, which control for time speaking (Gillespie et al., 2014; Hostetter & Alibali, 2007).

The goal of the current study was to identify whether gesture plays a role in lexical retrieval using a well-established method of eliciting and coding TOTs (Gollan & Brown, 2006). And if so, whether the role for gesture in lexical retrieval would be better explained by the lexical retrieval hypothesis, the cognitive load account, or the motor activation account. We first conducted a between-subjects picture-naming study where some participants were allowed to gesture, some had their hands restrained, and some were allowed to gesture but had their neck restrained. If participants who were allowed to gesture ultimately retrieved more target words and resolved more TOTs than participants whose gestures were inhibited, then gesture clearly affects lexical retrieval. If participants with their necks inhibited performed as well as those allowed to gesture, then gesture inhibition, not general movement inhibition, negatively affects lexical retrieval. In addition we considered the locus of gesture’s effect on lexical retrieval by investigating whether gesture inhibition reduces correct retrievals and increases TOTs, which

would indicate that gesture supports lemma selection, or whether gesture inhibition only reduces TOT resolutions, which would isolate the effect of gesture to the retrieval of the word form.

In order to distinguish between the three accounts of gesture's role in lexical retrieval, we analyzed the types of gestures produced by participants who were free to gesture. The strongest case in support of the lexical retrieval hypothesis would be evidence indicating that iconic gestures that were transparently related to the target word served as the best facilitators of lexical retrieval. Alternatively, if, as the cognitive load account suggests, speakers use gesture to offload the cognitive work of lexical search, then meaningful gestures (iconic, metaphoric, and deictic) that are not necessarily transparently related to the target word should support lexical retrieval just as well as semantically-related iconic gestures. Alternatively, if the motor activation account holds, then non-meaningful movements (beat gestures and self-adaptors) should also be related to successful lexical retrieval.

Experiment 1 was followed by a second study (Experiment 2) to examine whether gesture's role in lexical retrieval is shaped by individual differences in gesture use and/or in short-term memory.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Ninety female English-speaking monolingual undergraduates recruited from a women's college ($M = 19.82$, range: 18-22) received course credit or a small payment to complete a picture naming task in one of three conditions: gesture allowed, gesture inhibited, or neck inhibited. We obtained informed consent from all participants. Because TOT rates are affected by age (Burke et al., 1991) and vocabulary knowledge (Gollan & Brown, 2006), we adopted a matched participants design that individually matched participants in one condition to a participant in each of the other two conditions on age (within 6 months) and years of college education (within one semester). Ten additional participants were tested, but eliminated because of background factors that could have affected lexical retrieval abilities such as bilingualism ($n=4$), being significantly outside the undergraduate age range of 18-22 years ($n=3$), or because of video recording errors ($n=3$).

2.1.2. Materials

Participants completed a picture-naming task with 52 black and white pictures of objects with low-frequency names designed to elicit a high number of TOTs (Gollan & Brown, 2006; see Appendix).

2.1.3. Procedure

All participants were told that investigators were interested in how fast participants could name pictured items. Participants in the neck-inhibited and the gesture-inhibited conditions were additionally told that we were interested in how restraining different parts of the body could affect spoken language production. Participants in the gesture inhibited condition put their hands in a pair of gloves attached with Velcro to a wooden board placed on a table (e.g., Pine et al., 2007). In the neck-inhibited condition, participants wore a foam cervical collar with adjustable Velcro straps. Participants in all three conditions were told that a TOT is “that frustrating feeling you have when you are sure you know a word but cannot recall it at that particular moment,” and they were encouraged to report whenever they were experiencing a TOT during the picture-naming task. In addition, the experimenter asked about the presence of a TOT state whenever the participant hesitated or indicated, verbally or non-verbally, that they were searching for the correct word while naming the pictures (e.g., Beattie & Coughlin, 1999). The experimenter presented the pictures in a fixed order, one at a time in a binder and monitored the time of the trial using a watch to ensure that the trial did not exceed 30 seconds. If the participant was unable to retrieve the word within 30 seconds, the experimenter, who was present throughout the task, said the target word and followed up by asking the participants if they had been experiencing a TOT, if they knew the target word, and if the target word was the one they were trying to retrieve, following the procedures laid out in Gollan and Brown (2006). Participants were filmed for later coding (30fps), and they were told that they were being recorded in order to facilitate the transcription of their English responses. Experimenters were blind to directional hypotheses about TOTs and gesture/movement.

2.1.4. Coding

Following Gollan and Brown (2006), responses were classified into the following categories: (a) a correct retrieval, (b) a TOT for the target word that either ended in an unsuccessful retrieval of the target word or a self-resolved retrieval of the target word, (c) a negative TOT was a self-resolved or unsuccessful retrieval of a different word, e.g., *asteroid* for *comet*, (d) a failed retrieval of a word that the participant knew, and (e) a “don't know” was a

word that the participant did not know. We additionally calculated the proportion of self-resolved TOTs by dividing the number of self-resolved TOTs by the total number of TOTs.

A primary coder blind to hypotheses coded all of the gestures produced during the gesture allowed and the neck-inhibited conditions, classifying gestures into one of three categories: semantically-related iconic gestures (e.g., miming using an ax for the target picture of an ax), non-specific representational gestures which were gestures that were related to the target in a less iconically transparent way (e.g., a deictic point to the picture or a metaphorical “conduit” gesture; McNeill, 1992), and non-meaningful motor movements (e.g., beats and self adaptors). Only gestures that occurred from the time the picture was presented to the completed production of the target word or after 30 seconds, whichever came first, were coded. We coded the end of a gesture if the hands paused for at least one-tenth of a second (3 video frames) or if the gesture transitioned into a different gesture type (e.g., beat to iconic) or adopted a different handshape (e.g., beat with an open hand that switched to a beat with a closed fist). Gestures produced after successful retrieval or in between trials were not included in the analyses. We computed gesture rates by dividing the number of gestures by the number of English words produced (e.g., Nicoladis, Pika, & Marentette, 2009), including verbal fillers such as “uh” and “um” (see Clark & Fox Tree, 2002, for a discussion of the lexical status of verbal fillers) and the target words for the picture-naming task. A second coder, blind to hypotheses, coded the gestures produced by a randomly selected sample of thirty percent of the participants ($n=16$) in the gesture allowed and neck inhibited conditions for reliability (391 total gestures). Agreement between the two coders was high (total gestures: 85.38%, $r=.99$; semantically-related iconic gestures: 97.69%, $r=.95$; non-specific representational gestures: 90%, $r=.90$; non-meaningful motor movements: 83.85%, $r=.98$).

2.1.5. Notes on Mixed-Effects Models

We ran mixed effects logistic regressions with random intercepts for items and individual participants, and fixed effects of condition using the package *lme4* in the R Statistical Software (Version 3.4.2). All mixed effects logistic regressions were carried out with the function “*glmer()*” which specifies maximum likelihood. Reported p -values are based on asymptotic Wald tests. The fixed factor of condition was dummy coded (gesture allowed, gesture inhibited, neck inhibited; gesture allowed was the reference category).

2.2. Results

We ran several mixed effects logistic regressions following the specifications in the preceding Notes on Mixed-Effects Models section to ensure that the three experimental groups did not differ in the total number of English words produced and that the two groups who were allowed to gesture did not differ in the total number of gestures produced (see Table 1 for descriptive statistics).

Table 1. *Means and Standard Deviations of the English and Gesture Use Across Groups in Experiment 1*

	# English Words (TOT)	# English Words (not in TOT)	Semantically Related Iconic Gesture Rate	Non-specific Representational gesture rate	Non-meaningful motor movements
Group	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Gesture Allowed	12.40 (6.59)	3.70 (1.59)	.02 (.03)	.05 (.07)	.24 (.09)
Gesture Inhibited	14.14 (10.54)	3.59 (2.76)	-	-	-
Neck Inhibited	13.81 (9.04)	4.01 (1.53)	.04 (.05)	.06 (.07)	.13 (.09)

Note: Bolded values are significant at $p=.04$ when compared to the Gesture Allowed condition;
all other effects were non-significant (see Table A in supplementary materials)

Table 2. Means & Standard Deviations Comparing Groups for Each Response Type in Experiment 1

Group	Correct Retrieval	TOT	Proportion of Self Resolved TOTs	Negative TOT	Unsuccessful Retrieval	Don't Know
	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>
Gesture Allowed	38.80(5.52)	10.27(4.23)	.75(.16)	0.10(0.31)	2.07(1.44)	0.77(1.01)
Gesture Inhibited	40.07(5.39)	9.37(3.88)	.61(.27)	0.03(0.18)	1.93(1.93)	0.60(1.04)
Neck Inhibited	39.20(5.24)	9.53(3.97)	.70(.23)	0.10(0.31)	2.40(1.69)	0.77(0.73)

Note: Bolded values are significant at $p=.05$ when compared to the Gesture Allowed condition; all other effects were non-significant (See Table 3)

Participants in the gesture-inhibited condition and the neck-inhibited condition did not produce more English words than those in the gesture-allowed condition while in a TOT and while not in a TOT (see Table 1 for descriptive statistics, and Table A in Supplementary Materials for fixed and random effects for all variables). We applied the same statistical approach to compare gesture use between the two groups who were allowed to gesture—gesture allowed and neck-inhibited—and found that they did not differ significantly in their overall gesture rate, their non-specific iconic gesture rates, or their rate of non-meaningful motor movements, but they did differ significantly in their semantically-related iconic gesture rates ($\beta=0.008$, S.E. =0.004, $p=.04$; 95%CI [0.001; 0.016]).

We ran mixed effects logistic regressions with each of the different response types for the picture-naming task as the dependent variable (see Table 2 for descriptive statistics and Table 3 for the model summaries and the results for the central response types, and Table B in the Supplemental Materials for results related to peripheral response types). We observed no effect of experimental group on correct retrievals, TOTs, negative TOTS, unsuccessful retrievals, and unknown words. These findings indicate that gesture plays no role in overall lexical retrieval success. However, we observed a small, but significant effect of group in self-resolved TOTs: participants who were free to gesture resolved more of their TOTs than those who were inhibited from gesturing (Fig. 1) paralleling previous findings (Beattie & Coughlin, 1999; Pine et al., 2007). We did not observe the same effect of neck inhibition on self-resolved TOTs.

Table 3. Fixed and random effects for the central response outcomes on the picture naming task in Experiment 1. Table generated using the `tab_model` function reporting untransformed data in `sjplot` (Lüdtke, 2019)

Predictors	Correct Retrievals			TOTS			Self-Resolved TOTs		
	Log-Odds	CI	p	Log-Odds	CI	p	Log-Odds	CI	p
(Intercept)	1.52	1.04 – 2.00	<0.001	-1.75	-2.12 – -1.39	<0.001	1.38	0.90 – 1.87	<0.001
Gesture Inhibited	0.19	-0.20 – 0.57	0.340	-0.13	-0.43 – 0.17	0.381	-0.58	-1.18 – 0.01	0.054
Neck Inhibited	0.06	-0.33 – 0.44	0.773	-0.11	-0.41 – 0.19	0.479	-0.30	-0.89 – 0.30	0.330
Random Effects									
σ^2	3.29			3.29			3.29		
τ_{00}	0.44 ID.number			0.20 ID.number			0.71 ID.number		
	2.05 Item.Number			1.15 Item.Number			0.51 Item.Number		
ICC	0.43			0.29			0.27		
N	90 ID.number			90 ID.number			90 ID.number		
	52 Item.Number			52 Item.Number			51 Item.Number		
Observations	4680			4680			875		
Marginal R ² / Conditional R ²	0.001 / 0.431			0.001 / 0.291			0.013 / 0.280		

Note: `glmer(DV ~ Gesture.Condition + (1 | ID.number) + (1 | Item.Number), data = df, family = binomial, control = glmerControl(optimizer = "bobyqa"), nAGQ = 1)`

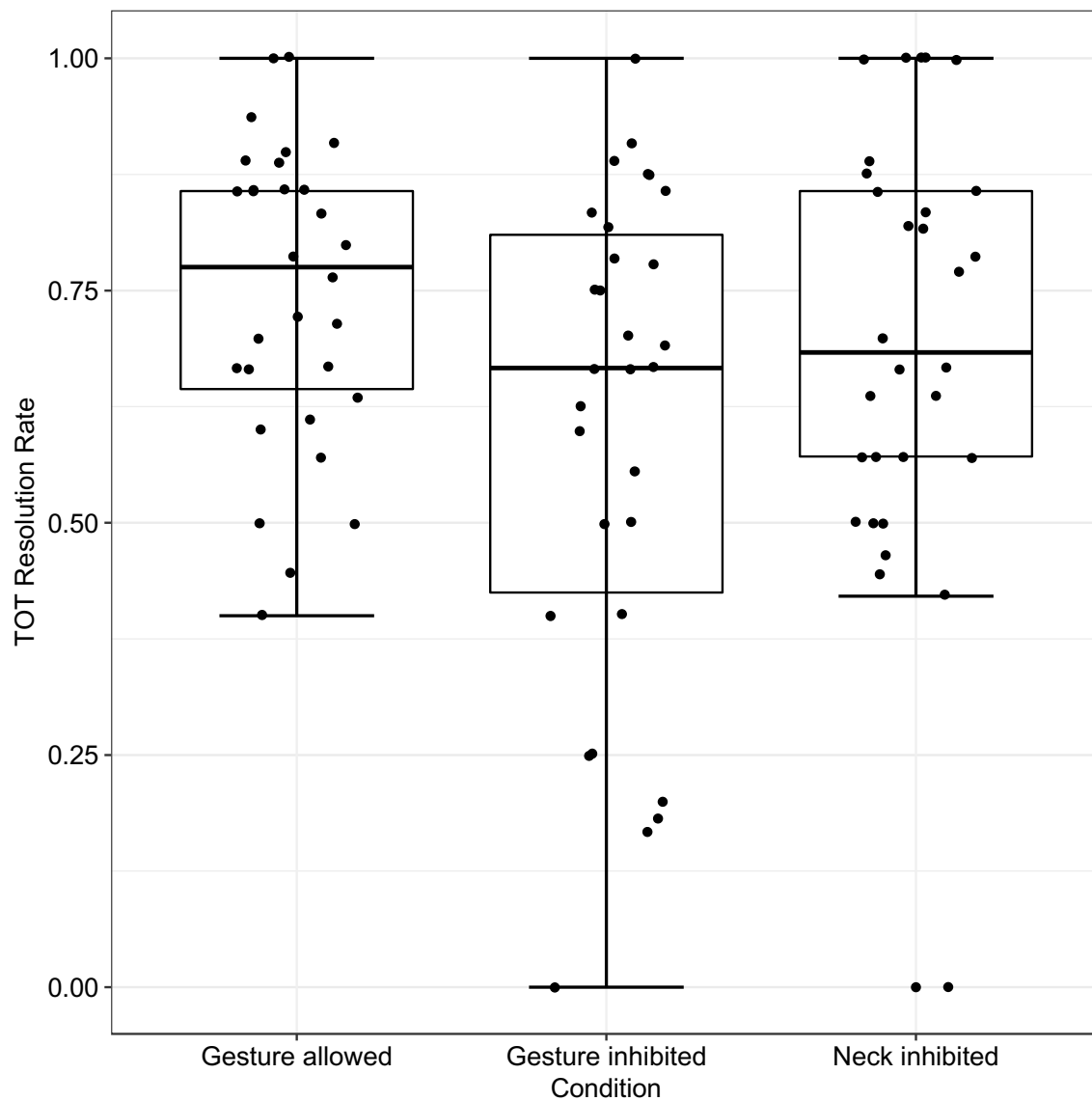


Figure 1. Boxplots depicting the proportion of self-resolved TOTs in each condition. Individual participants are depicted with a dot. We observed significantly greater variability in TOT resolution in the gesture-inhibited group compared to the gesture-allowed group. Plot was created using ggplot 2 for the R programming environment (Wikham, 2016).

If gesture supports the resolution of TOTs, then for those participants who were allowed to gesture (i.e., the gesture-allowed and neck-inhibited conditions), we should see higher gesture rates when they resolved a TOT compared to when they did not resolve their TOT within the 30-second window allowed for retrieval. Fifty-two of the sixty participants in the gesture-allowed and neck-inhibited groups gestured while experiencing a TOT. For these participants, we ran paired *t*-tests comparing rates of the three gesture types during successful vs. unsuccessful TOT resolution, combining participants from the gesture-allowed and neck-inhibited conditions. We observed a significantly higher rate of semantically-related iconic gestures and a significantly higher rate of non-specific representational gestures when the TOT was ultimately resolved compared to when it was not (see Table 4). Because participants in the neck-inhibited condition produced significantly more semantically-related gestures than those in the gesture-allowed condition, we examined these relationships within condition. When we do so, we see that the participants in the neck-inhibited condition primarily carry the observed effect of semantically-related gestures. We observed a marginally higher rate of non-specific representational gestures during resolved compared to unresolved TOTs among the participants in the gesture-allowed condition. We did not observe any significant difference in the rate of non-meaningful motor movements within any of the conditions, even though these were the most frequent hand movements observed (see Table 1).

Table 4. *Descriptive and inferential statistics for Experiment 1 and Experiment 2 examining rates of different gesture types when TOTs were resolved or unresolved*

		Semantically Related Iconic Gesture Rate <i>M(SD)</i>	Non-specific representational Gesture Rate <i>M(SD)</i>	Non-meaningful movement Rate <i>M(SD)</i>
	<i>n</i>			
Experiment 1				
All Gesturers	52			
Resolved		.030 (.045)	.060 (.082)	.174 (.132)
Unresolved		.017 (.024)	.027 (.032)	.175 (.239)
<i>t</i>		2.105	2.99	0.047
<i>p</i>		.040	.004	.962
95% CI Lower		-.026	-.054	-.059
95% CI Upper		-.0006	-.011	.062
Gesture-allowed	28			
Resolved		.023(.037)	.054(.089)	.214(.130)
Unresolved		.022(.030)	.027(.034)	.240(.306)
<i>t</i>		0.171	1.660	.510
<i>p</i>		.865	.109	.614
95% CI Lower		-.017	-.059	-.080
95% CI Upper		.014	.006	.133
Neck-inhibited	24			
Resolved		.039(.053)	.067(.074)	.127(.120)
Unresolved		.011(.014)	.027(.030)	.100(.076)
<i>t</i>		2.796	2.715	1.134
<i>p</i>		.010	.012	.268
95% CI Lower		-.047	-.069	-.078
95% CI Upper		-.007	-.009	.023

Experiment 2

Gesture-allowed 32

Resolved	.046(.073)	.103(.163)	.134(.148)
Unresolved	.016(.020)	.049(.068)	.174(.139)
<i>t</i>	2.418	1.875	1.275
<i>p</i>	.022	.070	.212
95% CI Lower	-.545	-.112	-.024
95% CI Upper	-.005	.005	.104

Note: The bolded *p*-values highlight significant results

The effect of gesture inhibition on TOT resolution was not uniform across participants. Specifically, the gesture-inhibited group exhibited significantly greater variability than the gesture-allowed group in their ability to successfully resolve their TOTs (Levene's test for equality of variance, $F(1, 58) = 4.073$, $p = .048$; See Fig. 1): some participants were unaffected by gesture inhibition, while others struggled greatly when they were not allowed to gesture. The same difference in variability was not observed between the neck-inhibited and gesture-allowed groups (Levene's test for equality of variance, $F(1, 58) = 2.564$, $p = .115$; see Fig. 1).

2.3. Discussion

In the current experimental paradigm, designed to elicit frequent TOT experiences, a speaker's ability to produce gestures affected the likelihood that they would resolve a TOT but did not affect the overall number of immediate successful retrievals or number of TOTs. Thus, with respect to lexical retrieval, gesture inhibition does not cause a speaker to be better or worse at selecting the correct lemma. Only once in a TOT, when the speaker was faced with the challenge of selecting the right word form, did we observe an effect of gesture inhibition. Crucially, participants who had their necks inhibited did not exhibit the same difficulty in resolving their TOTs, indicating that the inhibition of gesture, not body movement, impaired TOT resolution. That gesture affects only the retrieval of the word form is compatible with all three hypotheses for the role of gesture in lexical retrieval, but calls for a refinement of the

lexical retrieval hypothesis to narrow the locus of the effect of gesture to phonological form retrieval.

The production of non-meaningful motor movements during a TOT does not seem to facilitate TOT resolution because such gestures appeared equally often when TOTs were resolved and when they were not resolved. We see some modest correlational evidence that the production of representational gestures might support TOT resolution in that they appeared more frequently when TOTs were resolved, although this effect was primarily seen in the neck-inhibited condition. This pattern in the data rules out the motor activation account as a viable explanation for the role of gesture in TOT resolution. Given that we found differences between conditions in the patterns of gesture use during a TOT, we cannot definitely rule out the lexical retrieval hypothesis' proposal that gestures with a high degree of semantic overlap should best facilitate lexical retrieval; we did not see this effect in the gesture-allowed condition, but we did see the effect in the neck-inhibited condition.

Finally, we unexpectedly observed greater variability in TOT resolution among participants in the gesture-inhibited condition compared to other two conditions. This variability could be attributed to one of two factors. First, individuals could vary in their gesture use as a function of individual differences in experience, and it is possible that high gesture users, who are accustomed to using gesture, are more negatively affected in the inhibition condition than low gesture users. Alternatively, individuals may vary in lexical retrieval skill, and, as such, may vary in the degree to which they need to rely on gesture to lift the cognitive burden when searching for a word form. That is, everyone may gesture, but only some depend on gesture to offload the difficult task of lexical retrieval, particularly of low frequency words that often lead to TOT states. In Experiment 2 we investigated individual differences that may contribute to the variability of the effect of gesture inhibition on lexical retrieval.

3. Experiment 2

Each of the three theoretical accounts for the role of gesture in lexical retrieval considers how working memory and/or short-term memory functions in this relationship (investigating both spatial and verbal memory). Spatial memory has been linked to gesture use in a variety of studies. Participants gesture more when gesturing about objects from memory (Morsella & Krauss, 2004; Wesp, Hesse, & Keutmann, 2001) and when gesturing about complex images. Both of these circumstances are presumed to place demands on spatial memory (Morsella &

Krauss, 2004). Because gesture production increases when an individual's spatial memory is taxed, it may be the case that individuals with weaker spatial memory may gesture more overall compared to those with stronger spatial memory skill. In this situation, gesture can help maintain the mental image during lexical retrieval, and those with weaker spatial skills need more help from gesture (Wesp et al., 2001). In support of such an account, Chu, Meyer, Foulkes, and Kita, (2014) found a strong negative relationship between spatial memory (both working and short-term memory) and spontaneous use of representational gestures during a task where participants defined phrases and described how they would handle social dilemmas; those with weaker spatial memory tended to use more representational gestures. Alternatively, Hostetter and Alibali (2007) suggested that strong spatial skills allow for better image generation, rather than maintenance, which in turn supports the production of gestures. Supporting this account, they found that strong visualization skills positively correlated with representational gesture use, specifically for those participants who also had weaker phonemic fluency skills, as measured by a task asking participants to recall as many words as they could that began with a specific letter. Yet, countering the above findings, a third study found that spatial working memory and spatial short-term memory did not predict any variance in participants' gesture use, iconic or otherwise, during a narrative task (Smithson & Nicoladis, 2013).

Verbal memory may be another variable underpinning gesture's role in lexical retrieval. There is modest evidence of a relationship between verbal working memory and speech production in terms of picture naming (Belke, 2008), verbal fluency (Daneman, 1991), and sentence production (Slevc, 2011). In a lexical retrieval task, people with stronger verbal memory may not experience the same cognitive burden as those with weaker verbal memory. Thus, those with weaker verbal memory may use gesture, particularly representational gesture, to help maintain semantic information during the search for the word form. Some research lends support to the association between low verbal abilities and high gesture use. Hostetter and Alibali (2007) found an interactive effect of high visual-spatial skills and low verbal ability, as assessed by phonemic fluency, on representational gesture use. Further, Wagner et al., (2004) observed that participants who were allowed to gesture while solving a math problem were better able to hold on to a set of to-be-remembered numbers compared to when they were not allowed to gesture.

These observed relations between gesture and verbal memory have been further bolstered by recent work on individual differences that explicitly tested the relationship between individuals' gesture use and their performance on different verbal working memory tasks. Using a composite measure of listening span and a "subtract two" span, Gillespie et al. (2014) found a negative relationship between verbal working memory and overall gesture and iconic gesture rates. Specifically, they observed lower gesture rates for participants with higher verbal working memory scores. Relatedly, Smithson and Nicoladis (2012) found that verbal short-term memory as measured by digit span, but not listening span (a working memory measure), reliably predicted iconic gesture use: weaker verbal short-term memory was associated with higher iconic gesture production during a narrative task. Interestingly, Gillespie et al. (2014) found no relationship between verbal fluency (semantic or phonemic) and iconic gesture production, leading them to conclude that the relationship between gesture and verbal memory is independent of lexical retrieval and instead reflects the demands of speech planning. Countering the evidence of a relationship among gesture use, verbal memory, and lexical retrieval, however, is the finding of no relationship among iconic gesture use, digit span, and picture naming latencies from a large-scale study ($N = 129$; Chu et al., 2014).

No study to date has specifically examined the effect of individual differences in spatial and verbal memory skills on gesture use during a lexical retrieval task, and on lexical retrieval when gesture is inhibited. With Experiment 2, we fill this gap in the literature by conducting a within-subjects study where all participants completed picture-naming tasks while allowed to gesture and while inhibited from gesturing. Crucially we examined participants' gesture rates while in the gesture-allowed condition, alongside their performance on spatial and verbal short-term memory tasks. While most studies looking at verbal or spatial memory and gesture use have tested only working memory, we specifically looked at short-term memory because when previous studies have included short-term memory measures (spatial or verbal) along alongside analogous measures of working memory, it was frequently the more powerful predictor of gesture-use (Chu, et al., 2014; Smithson & Nicoladis, 2012). In addition, working memory involves the short-term storage of information while simultaneously processing a different set of information; short-term memory captures the storage component of working memory without the additional processing demand, allowing us to isolate the effects of short-term storage capacity on lexical retrieval.

We considered three alternative hypotheses to explain the effect of gesture inhibition on self-resolved TOTs observed in Experiment 1. One possibility is that participants who gesture more during lexical retrieval will be more negatively affected when they are prevented from gesturing, regardless of their memory span scores, indicating that gesture may not have an off-loading function; rather, individual habits of gesturing may shape how much gesture inhibition affects lexical retrieval. Alternatively, participants' memory span scores could predict their ability to resolve TOTs in either the gesture-allowed or the gesture-inhibited conditions, providing some insight into the cognitive load associated with lexical retrieval. Specifically, if those with weaker spatial span scores resolve fewer TOTs in the gesture-inhibited condition, then these participants need to rely more on gesture for image maintenance during lexical retrieval. Alternatively, if we observe a strong relationship between verbal short-term memory and TOT resolution in the gesture-inhibited condition, then those with weaker verbal memory spans may find the retrieval of the phonological form more difficult than those with stronger verbal memory spans, and therefore they need gesture to reduce the cognitive load associated with word form retrieval. We consider that we may only observe effects of memory span in the gesture-inhibited condition where we prevent participants from off-loading some of the cognitive work of lexical retrieval to gesture. It could be that when everyone is allowed to gesture, effects of memory span on lexical retrieval disappear because gesture is effective in reducing cognitive load and in supporting word-form retrieval.

3.1. Method

3.1.1. Participants

Thirty-four English-speaking monolinguals recruited from a women's college campus ($M_{\text{age}} = 19.32$, range 18-22; 33 female) participated in this study for credit in a psychology course. Six additional participants were excluded for being bilingual ($n=1$), for falling outside of our age criteria ($n=2$), or due to video recording error ($n=4$). All participants consented to participate in the study.

3.1.2. Materials

We developed two sets of stimuli: one included 52 pictures of the same words elicited in Experiment 1 taken from Gollan and Brown (2006; frequency: $M=2.92$, $SD=4.26$, Baayen, Piepenbrock, & Gulikers, 1995; length: $M=6.52$, syllable length: $M=1.98$, Balota et al, 2007), and the other included a second set of 52 pictures of imageable low-frequency words (frequency

$M = 5.31$, $SD = 5.3$, Baayen et al, 1995; length: $M = 6.706$, syllable length: $M = 2.14$, Balota et al, 2007, see Appendix), primarily selected from stimuli used in other studies that investigated TOTs (e.g., Beattie & Coughlin, 1998; Frick-Horbury & Guttentag, 1998) and further piloted to confirm they elicited a high number of TOTs. Images were presented on a 13-inch MacBook G5 laptop in an automated PowerPoint slideshow.

3.1.3. Procedure

Participants named one set of pictures while allowed to gesture and a second set while inhibited from gesturing. Gesture condition and picture set were counterbalanced across participants. The picture naming tasks followed the same procedure as Experiment 1 with the exception that the 30-second window to respond was automatically controlled in the PowerPoint slide show. Following the picture-naming task, participants were tested on spatial short-term memory (Corsi Block Tapping Test; Milner, 1971) and verbal short-term memory (Wechsler Digit Span Task; Wechsler, 1949), order counterbalanced. The experimenters were blind to the hypotheses.

3.1.4. Coding

Verbal responses in the picture-naming tasks were coded in the same fashion as Experiment 1 by a coder blind to hypotheses, but not to condition. We also coded all hand movements produced during the picture-naming task in the gesture allowed condition and computed gesture rates following the procedures outlined in Experiment 1. Inter-rater agreement between the primary coder who coded all participants' gestures, and secondary coder, who coded 30% of the participants ($n = 17$; 325 gestures total), was high (total gestures: 87.58%, $r = .94$; semantically-related iconic gestures: 96.89%, $r = .90$; non-specific representational gestures: 89.40%, $r = .95$; non-meaningful motor movements: 87.58%, $r = .85$). Both gesture coders were blind to hypotheses.

3.1.5. Notes on Mixed-Effects Models

We adopted a similar mixed effects logistic regression model procedure using the same packages and software outlined in Experiment 1 to examine the effects of gesture condition on the response categories. We included items and participants as random effects and condition as a fixed effect. Condition had only two levels, dummy coded as "gesture allowed" and "gesture inhibited" with "gesture allowed" as the reference category.

When exploring the effects individual differences in gesture use, spatial span, and digit span on TOT resolution, we ran a series of mixed effects logistic regressions with items and participants as random effects and representational gesture rate, spatial span, and digit span as fixed effects.

3.2. Results and Discussion

We ran mixed effects logistic regressions following the procedures outlined in the above section (Notes on Mixed-Effects Models) with each of the different response categories for the picture-naming task as the dependent variable (see Table 5 for descriptive statistics and Table C in the Supplementary Materials for fixed and random effects). We found no effect of condition in any response category, including the ability to resolve TOTs. The fact that we did not see a difference in TOT resolution rates between conditions in this within-subjects study is unsurprising given that individual differences remain constant across conditions and the observed between-subjects condition effect on TOT resolution in Experiment 1 was small.

As in Experiment 1, we examined the rates of different gesture types produced by the 32 participants who gestured while having a TOT (see Table 4). In contrast to the pattern we observed in the gesture-allowed condition in Experiment 1, we observed a significantly higher rate of semantically-related iconic gestures and a marginally higher rate of non-specific representational gestures when the TOT was ultimately resolved compared to when it was not. As we found in Experiment 1, this effect was not observed for non-meaningful motor movements. Thus, representational gestures, semantically-related or otherwise, tend to appear more frequently with resolved TOTs than with unresolved TOTs, providing further evidence for the specific role of representational gestures in lexical retrieval.

Table 5. *Means, Standard Deviations, and Statistics Comparing Conditions for Each Response Type in Experiment 2*

	Correct Retrieval	TOT	Proportion of Self Resolved TOTs	Negative TOT	Unsuccessful Retrieval	Don't Know
Condition	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>
Gesture Allowed	39.09(5.28)	10.47(4.23)	.70(.18)	0.06(0.34)	1.76(1.56)	0.62(0.99)
Gesture Inhibited	39.41(6.04)	10.06(4.40)	.69(.21)	0.19(0.39)	1.79(1.70)	0.53(1.08)

To explore the effects of individual differences in gesture use and short-term memory on TOT resolution collapsed across both conditions, we first ran a mixed effects logistic regression with participant and item as random factors and gesture rate for all representational gestures (semantically-related and non-specific), spatial span, and digit span as fixed effects. All predictor variables were z-transformed for all analyses. None of the variables significantly predicted TOT resolution (see Table D in the Supplementary Materials), although spatial span had a marginally significant effect ($\beta = 0.225$, S.E. = 0.119, $z = 1.896$, $p = .058$; 95%CI [-0.008, 0.457]); those with stronger spatial spans had numerically, but non-significantly, higher TOT resolution rates ($r_s = .31$, $p = .08$).

Of theoretical interest, however, was the role of these factors in predicting TOT resolution when gesture was inhibited, given that the results of Experiment 1 showed the greatest variability in this condition. We conducted a second mixed effects logistic regression predicting TOT resolution by examining the role of each factor on TOT resolution rates while gesture was inhibited. In order to account for participants' general ability to resolve TOTs, we included participants' TOT resolution rate (the proportion of self-resolved TOTs) while they were in the gesture-allowed condition in the model. Thus, in the model we treated item as a random factor¹ and TOT resolution rates while gesturing, gesture rate for all representational gestures, spatial span, and digit span as fixed effects. All predictor variables were z-transformed. Table 6 presents all random and fixed effects. As expected, increased ability to resolve TOTs when allowed to gesture predicted the likelihood of TOT resolution when gesture was inhibited. Crucially, digit span was the only other significant predictor of the likelihood of TOT resolution in the gesture inhibited condition, such that those with weaker verbal short-term memory were less able to find a target word once in a TOT if they were not allowed to gesture. Representational gesture rate and spatial span scores did not play the same role. Thus, after controlling for an individual's overall ability to resolve TOTs, only individual differences in verbal short-term memory, not spatial short-term memory, and not in an individual's propensity to produce representational gestures, predicted the ability to resolve TOTs when the cognitive load associated with lexical retrieval could not be offloaded to gesture (see Fig. 2)

¹ The inclusion of participant as a random factor led to issues of singularity in the model, so following Barr et al (2013), we removed participant as a random factor.

We confirmed the significant effect of digit span on TOT resolution when gesture was inhibited using a log-likelihood test to compare a model containing digit span to a model excluding it to determine whether its inclusion was justified (Barr, Levy, Scheepers, & Tily, 2013). The full model significantly outperformed the model without digit span ($\chi^2(1) = 4.32, p = .038$; see Table 6).

Table 6. *Fixed and random effects for self-resolved TOTs in the Gesture Inhibited condition in Experiment 2. Table generated using the tab_model function reporting untransformed data in sjplot (Lüdtke, 2019)*

<i>Predictors</i>	Model 1			Model 2		
	<i>Log-Odds</i>	<i>CI</i>	<i>p</i>	<i>Log-Odds</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.91	0.60 – 1.22	<0.001	0.90	0.60 – 1.21	<0.001
TOT Resolution	0.29	0.06 – 0.53	0.015	0.28	0.04 – 0.51	0.021
Gesture Allowed						
Representational Gesture	0.14	- 0.09– 0.38	0.233	0.12	- 0.11– 0.35	0.319
Spatial Span	0.17	- 0.09– 0.44	0.202	0.23	-0.0 – 0.49	0.079
Digit Span	0.27	0.01 – 0.54	0.041			
Random Effects						
σ^2	3.29			3.29		
τ_{00}	0.26 Item			0.21 Item		
ICC	0.07			0.06		
N	87 Item			87 Item		
Observations	342			342		
Marginal R ² / Conditional R ²	0.073 / 0.141			0.055 / 0.112		
AIC	421			423.2		

Note: glmer(TOT_Resolution ~ TOT_Resolution_Gesture_Allowed

+Representational_Gesture + Spatial_Span + Digit_Span +(1|Item), data = df2, family =

binomial, control = glmerControl(optimizer = "bobyqa"), nAGQ = 1). We excluded participant as a random factor because of issues of singularity in the model (Barr et al., 2013). Bolded p-values are significant.

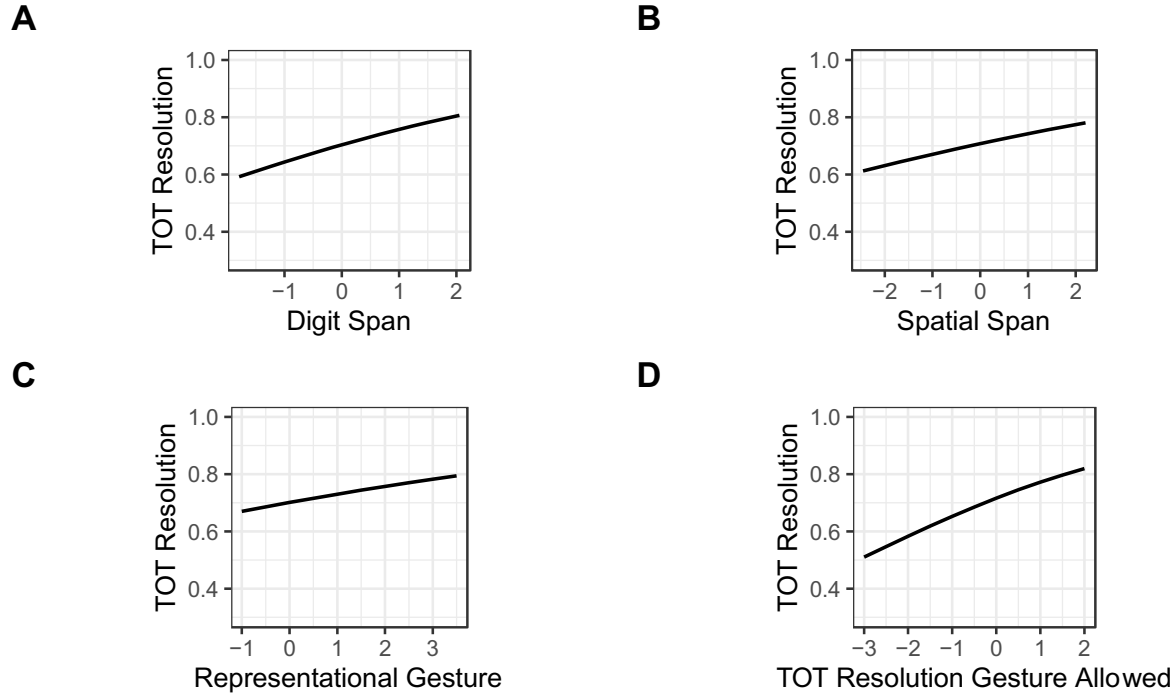


Figure 2. Predicted probability of resolving a TOT while gesture was inhibited as a function of the z-scored predictors (A) digit span, (B) spatial span, (C) representational gesture use, (D) TOT resolution ability while allowed to gesture. Bands around the regression lines indicate 95% CIs computed from SEMs. Plots were created using the `plot_model` function in `sjplot` (Lüdtke, 2019).

We ran the identical analyses predicting TOT resolution for the gesture-allowed condition. Only ability to resolve TOTs when gesture was inhibited predicted TOT resolution when gesture was allowed ($\beta = 0.33$, S.E. = 0.153, $p = .023$; 95%CI [0.071, 0.590]; see Table E in Supplementary Materials for the full model). The remaining variables, representational gesture rate, spatial span, and digit span, did not significantly predict TOT resolution. Thus, when participants were free to gesture, their memory spans, verbal or spatial, and their personal gesture rates played little role in predicting TOT resolution.

4. General Discussion

Inhibiting gesture did not affect overall lexical retrieval: participants retrieved the same number of words regardless of whether or not they were allowed to gesture in both Experiments. However, once lexical retrieval failed, as when in a TOT, inhibiting gesture seemed to make TOT resolution more difficult. As shown in Experiment 1, this result is specific to inhibiting movement of the hands but not other parts of the body. In Experiment 1, gesture inhibition affected individuals differently and some struggled more than others to resolve their TOTs (see Fig. 1). The within-subjects design of Experiment 2 allowed us to trace individual differences in successful TOT resolution while prevented from gesturing, after controlling for TOT resolution ability, to individual differences in verbal short-term memory, but not to differences in spatial short-term memory or gesture use. Those with weaker verbal short-term memory struggled to resolve their TOTs when they were not allowed to gesture, whereas those with larger verbal memory spans were unaffected when their gestures were inhibited. Thus, the effect of gesture inhibition on lexical retrieval is very specific and only seen when lexical retrieval is difficult, as in the case of a TOT, and primarily for individuals with weaker verbal short-term memory.

The detrimental effect of gesture inhibition on the resolution of TOTs is consistent with all three accounts of the role of gesture in lexical retrieval: the lexical retrieval hypothesis, the cognitive load account, and the motor activation account. However, examination of the gesture types produced during self-resolved TOTs likely rules out non-meaningful motor movements as a possible driving factor behind successful lexical retrieval (e.g., Ravizza, 2003). In both Experiments 1 and 2, semantically-related gestures and non-specific representational gestures appeared more often during self-resolved TOTs than during unresolved TOTs. In contrast, non-meaningful motor movements were not present to a greater degree during self-resolved TOTs. Rather, the numerical trend was towards more non-meaningful motor movements during unresolved compared to resolved TOTs. The study by Ravizza (2003) differs from our study in several ways that may account for our different findings: participants had to retrieve words from definitions instead of during a picture naming task, TOTs were elicited in an initial round of testing and then those items were presented a second time during the experimental manipulation (tapping vs. no tapping). Participants were instructed to type not speak any resolved TOTs. Self-

paced tapping may specifically support the typed retrieval of the form after extended exposure but not immediate TOT resolution during verbal picture naming tasks. Thus, it is unlikely that spontaneously generated motor movements prime lexical retrieval via activating neural areas common for both language and motor production.

The pattern of gesture use during resolved and unresolved TOTs in Experiments 1 and 2 is consistent with both the lexical retrieval hypothesis and the cognitive load account of the relationship between gesture and lexical retrieval, but the effect of gesture inhibition on lexical retrieval and the relationships to short-term memory help refine these accounts. First, the lexical retrieval hypothesis makes no commitment as to whether gesture cross-modally primes lemma selection or word form selection (Krauss & Hadar, 1999, Morsella & Krauss, 2004). Our data clearly isolate the effects of gesture on retrieving the phonological form of the word. The only effect of inhibition we observed was on TOT resolution rates, not on the number of correct retrievals or the number of TOTs. Thus, if we accept the lexical retrieval hypothesis' model of cross-modal priming, we have to exclude an effect at the lemma level. Second, we observed no effect of spatial short-term memory on TOT resolution. The previous study by Morsella and Krauss (2004) that showed an effect of spatial memory on lexical retrieval differed significantly from our study in that spatial memory was taxed by asking participants to describe complex shapes from memory while allowed to gesture or while inhibited from gesturing. In contrast, we used a standardized measure of spatial span to tap an individual's spatial memory storage capacity, rather than taxing spatial memory via the experimental task itself. However, our task required lexical retrieval of pictured objects, which likely placed little demand on spatial memory in comparison to tasks where images or videos with spatial content had to be remembered and then described from memory (e.g., Hostetter & Alibali, 2007; Morsella & Krauss, 2004).

The cognitive load account does not view gesture's role as one of necessarily priming the word form, but rather as a mechanism to reduce the cognitive load of any difficult task (Goldin-Meadow, Nussbaum, Kelly, & Wagner, 2001). In our study, the difficult task for participants was finding a word form while experiencing a TOT. Thus, the effect of gesture on lexical retrieval in our study is arguably an effect of gesture helping to relieve the cognitive load of TOT

resolution. This mechanism is not necessarily specific to lexical retrieval but could be applied across individuals for whatever task they find difficult.

Most studies investigating the cognitive load account have used measures of verbal working memory to capture the difficulty associated with tasks that gesture may support. Our results expand these findings by showing that gesture can reduce the cognitive load placed specifically on verbal short-term memory during lexical retrieval. The presence of a greater number of representational gestures during self-resolved TOTs is an indicator that the cognitive load of lexical retrieval is being successfully offloaded to gesture. In this case, gestures that capture any related features of the referent may help maintain semantic information about the referent in short-term memory while the speaker engages in the search for the phonological form of the referent. TOTs occur when access to a word's phonology is weak (e.g. Brown, 1991), and verbal memory may play a unique role in resolving TOTs because a TOT is, in part, a failure to access a word's phonology (Levelt, 1989). Some evidence suggests a positive association between verbal working memory and lexical retrieval for L1 speakers (Belke, 2008; Daneman, 1991) and a robust relationship between verbal working memory and L2 language production (see Linck, Osthus, Koeth, & Bunting, 2014 for a meta-analysis). As such, it follows that those with weaker verbal memory would have increased difficulty accessing the phonological form of a target word when in a TOT (e.g., when retrieval is difficult). One possible explanation for our finding that digit span scores predict successful TOT resolution when gesture is inhibited is that TOT resolution places demands on the retrieval of phonological information independent of additional cognitive demands placed by language processing. Thus, without the cognitive load reduction provided by gesturing, those with weaker verbal short-term memory struggle to resolve their TOTs. Crucially, gesture seems to be effectively performing this off-loading function for those with weaker verbal short-term memory because we observed no effect of digit span on TOT resolution in the gesture allowed condition.

The relationship between verbal short-term memory and TOT resolution while gesture was inhibited seems to strongly suggest a role for gesture in alleviating some of the cognitive burden during lexical retrieval. However, our study cannot rule out the lexical retrieval hypothesis' suggestion of cross-modal priming as a mechanism. For example, some lexical

decision studies have observed robust priming effects of viewing iconic gestures: participants were faster to identify a word when the word had been preceded by a video showing a semantically-related gesture compared to when they viewed a semantically unrelated gesture (So, Yi-Feng, Yap, Kheng, & Yap, 2013; Yap, So, Yap, Tan, & Teoh, 2011). It is possible that self-generated gestures could perform the same priming function for lexical retrieval, although the cross-modal priming effect would be motorically or somatosensorily, rather than visually driven. In addition, any cross-modal priming mechanism would have to consider how non-specific representational gestures in addition to semantically-related representational gestures can prime TOT resolution.

Studies that have challenged the role of gesture in lexical retrieval have relied on measures of verbal fluency to assess lexical retrieval ability, and have correlated verbal fluency scores with rates of gesture use while narrating a story from memory rather than with rates of gestures produced during lexical retrieval itself. In contrast, we examined gestures produced while retrieving (or attempting to retrieve) low frequency lexical items. During a narration task, gesture may be functioning at multiple levels, such as aiding the planning and execution of a narrative, as well as facilitating lexical retrieval. In our task, gesture was primarily produced during failed lexical retrieval (i.e., when in a TOT), and thus gesture was more likely tied to the task of retrieving the phonological form of low-frequency words, and not to supporting the organization of the discourse. Crucial to reiterate here, is that gesture rates did not predict the probability that participants would resolve a TOT when allowed to gesture. Rather, we observed a pattern of higher representational gesture rates with successful, as opposed to unsuccessful resolutions in both experiments. Taken together, these findings indicate that gesture plays a role not only in verbal planning, as suggested by the Information Packaging Hypothesis (Kita, 2000), but also in lexical retrieval.

To date, the role of gesture in lexical retrieval has been decidedly inconclusive, likely because of several factors, including methodological variability across studies and the limited attention paid to individual differences in lexical retrieval abilities. Our study attempts to overcome some of these limitations by relating gestures produced during a lexical retrieval task to performance on the same task (Experiments 1 & 2), by matching participants on factors

known to affect lexical retrieval, such as age and education level (Experiment 1), and by specifically investigating individual differences in factors hypothesized to underpin the relationship between gesture and lexical retrieval (Experiment 2). In doing so, we find that representational gestures support the retrieval of the word form and that the effect of gesture depends on the individual differences in verbal short-term memory span. Notably the effects we observe were modest and merit replication with a more diverse sample of participants and perhaps with different methods of eliciting lexical retrieval, such as recalling words from definitions instead of picture naming given that the former may place more demands on memory and elicit more gestures (Morsella & Krauss, 2004; Wesp, et al., 2001).

In sum, we find that while experiencing a TOT, some speakers' ultimate successful retrieval of a word form can be disrupted by inhibiting their ability to gesture. Representational gestures, semantically-related or otherwise, appear more frequently during TOTs that have been successfully resolved. Most importantly, speakers' reliance on gesture depends on how much they need it; speakers with weaker verbal short-term memory may find getting out of a TOT more difficult than those with stronger verbal short-term memory and as such, they may need to offload some of the cognitive work of lexical search to gesture. When speakers were allowed to gesture, we saw no effects of short-term memory ability on lexical retrieval success. Thus, in the case of lexical retrieval, gesture is a great equalizer.

Acknowledgements

We thank Masa Dikanovic, Pamela Doig, Sammantha Grossmith, Giselle Lehman, and Sara Mejia for research assistance.

Funding: This work was supported by the Wellesley College Summer Research program, a faculty award from the Wellesley College Provosts' office (J.E.P), a Miriam Cabot Putnam Memorial Fellowship from the Radcliffe Institute for Advanced Study (J.E.P), a grant from the James S. McDonnell foundation (J.E.P), and grants from the National Institutes of Health (R01 HD047736 and R01 DC010997 to K.E.).

Conflicts of Interest: none

References

- Baayen, R., Piepenbrock, R., & Gulikers, L. (1995) CELEX2 LDC96L14. Web Download. Philadelphia: *Linguistic Data Consortium*.
- Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445-459.
- 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.
- 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.
- Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review*, 15(2), 357-363.
- Brown, A. S. (1991). A review of the tip of the tongue phenomenon. *Psychological Bulletin*, 109, 204-223.
- Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, 143(2), 694-709.
- Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73-111.
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, 27(4), 594-610.
- Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, 20(6), 445-464.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2), 283-292.

- 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), 43-62.
- Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition*, 132(2), 174-180.
- Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
- Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: when more TOTs means better retrieval. *Journal of Experimental Psychology: General*, 135(3), 462-483.
- Gunderson, E. A., Spaepen, E., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2015). Gesture as a window onto children's number knowledge. *Cognition*, 144, 14-28.
- Holler, J., Turner, K., & Varciana, T. (2013). It's on the tip of my fingers: Co-speech gestures during lexical retrieval in different social contexts. *Language and Cognitive Processes*, 28(10), 1509-1518.
- Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, 7(1), 73-95.
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162-185). Cambridge, England: Cambridge University Press.
doi:10.1017/CBO9780511620850.011
- 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: Oxford University Press.
- Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861-883.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92(3), 350-371.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: 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*, 411-424.
- 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.
- 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.
- Ravizza, S. (2003). Movement and lexical access: Do noniconic gestures aid in retrieval? *Psychonomic Bulletin & Review*, 10(3), 610-615.
- Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1503.
- So, W. C., Low, A., Yap, D. F., Kheng, E., & Yap, M. (2013). Iconic gestures prime words: comparison of priming effects when gestures are presented alone and when they are accompanying speech. *Frontiers in Psychology*, 4, 779.
- Smithson, L., & Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 16(04), 934-944.
- Trofatter, C., Kontra, C., Beilock, S., & Goldin-Meadow, S. (2015). Gesturing has a larger impact on problem-solving than action, even when action is accompanied by words. *Language, Cognition and Neuroscience*, 30(3), 251-260.
- Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, 50(4), 395-407.
- Wechsler, D. (1949). *The Wechsler Intelligence Scale for Children*. New York: Psychological Corp.
- Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, 114(4), 591-600.
- Yap, D. F., So, W. C., Melvin Yap, J. M., Tan, Y. Q., & Teoh, R. L. S. (2011). Iconic gestures prime words. *Cognitive Science*, 35(1), 171-183.

Appendix

Target words used in Experiment 1

hive	chisels	walnut
peacock	ostrich	catapult
comet	grater	udder
microscope	braille	gyroscope
snorkel	well	mummy
megaphone	guillotine	plunger
pitcher	weathervane	hinge
mantaray	churn	harmonica
axe	unicycle	cleft chin
gazebo	thermos	metronome
hoe	carousel	handcuffs
baster	antlers	noose
castle	canteen	harp
eclipse	beaver	slingshot
bolt	goggles	Eiffel tower
wheelbarrow	boomerang	syringe
dynamite	easel	
compass	propeller	

Additional target words used in Experiment 2

mushroom	thermometer	bannister
tomato	owl	suitcase
mannequin	tiger	bagpipes
trellis	helicopter	wrench
origami	caterpillar	pediatrician
saddle	tambourine	hydrant
abacus	accordion	stethoscope
spool	javelin	bandana
turtle	shamrock	seesaw
canoe	donkey	iguana
record player	parachute	broom
crib	bleachers	chaps
gondola	umbrella	doorknob
pineapple	centaur	cactus
trumpet	chariot	ladder
tarot	anchor	binoculars
blouse	rolling pin	
clown	whistle	

Supplementary Materials

Table A. Fixed and random effects for language and gesture use on the picture naming task in Experiment 1. Table generated using the *tab_model* function reporting untransformed data in *sjplot* (Lüdtke, 2019)

Predictors	Number of English words in a TOT			Number of English words not in a TOT			Semantically Related Iconic Gesture Rate			Non-specific Representational Gesture Rate			Non-Meaningful Motor Movements		
	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
(Intercept)	11.56	8.12 – 15.01	<0.001	3.95	2.70 – 5.21	<0.001	0.01	0.00 – 0.01	0.044	0.03	0.02 – 0.04	<0.001	0.10	0.07 – 0.13	<0.001
Gesture inhibited	1.27	-3.33 – 5.87	0.589	0.02	-0.99–1.03	0.969									
Neck inhibited	1.45	-3.14 – 6.03	0.536	0.35	-0.66–1.36	0.496	0.01	0.00 – 0.02	0.036	-0.00	-0.02 – 0.02	0.909	-0.04	-0.08–0.00	0.075
Random Effects															
σ^2	146.08			34.68			0.00			0.02			0.04		
τ_{00}	64.60 ID.number			3.12 ID.number			0.00 ID.number			0.00 ID.number			0.01 ID.number		
	15.89 Item.Number			14.54 Item.Number			0.00 Item.Number			0.00 Item.Number			0.00 Item.Number		
ICC	0.36			0.34			0.06			0.05			0.13		
N	90 ID.number			90 ID.number			60 ID.number			60 ID.number			60 ID.number		
	51 Item.Number			52 Item.Number			52 Item.Number			52 Item.Number			52 Item.Number		
Observations	875			3805			3120			3120			3120		
Marginal R ² / Conditional R ²	0.002 / 0.356			0.000 / 0.338			0.004 / 0.065			0.000 / 0.049			0.007 / 0.131		

Table B. *Fixed and random effects for each peripheral response outcome on the picture naming task in Experiment 1. Table generated using the tab_model function reporting untransformed data in sjplot (Lüdtke, 2019)*

<i>Predictors</i>	Negative TOTs			Unsuccessful Retrieval			Unknown Word		
	<i>Log-Odds</i>	<i>CI</i>	<i>p</i>	<i>Log-Odds</i>	<i>CI</i>	<i>p</i>	<i>Log-Odds</i>	<i>CI</i>	<i>p</i>
(Intercept)	-8.60	-11.40 – -5.80	<0.001	-4.92	-5.80 – -4.04	<0.001	-8.92	-11.72 – -6.12	<0.001
Gesture Inhibited	-1.04	-3.70 – 1.62	0.444	-0.11	-0.63 – 0.40	0.663	-0.39	-1.34 – 0.56	0.422
Neck Inhibited	0.00	-2.06 – 2.06	1.000	0.18	-0.31 – 0.68	0.466	0.06	-0.84 – 0.97	0.889
Random Effects									
σ^2	3.29			3.29			3.29		
τ_{00}	3.67 ID.number			0.36 ID.number			1.11 ID.number		
	2.91 Item.Number			4.44 Item.Number			16.65 Item.Number		
ICC	0.67			0.59			0.84		
N	90 ID.number			90 ID.number			90 ID.number		
	52 Item.Number			52 Item.Number			52 Item.Number		
Observations	4680			4680			4680		
Marginal R ² / Conditional R ²	0.024 / 0.675			0.002 / 0.594			0.002 / 0.844		

Table C. Fixed and random effects for each response outcome on the picture naming task in Experiment 2. Table generated using the `tab_model` function reporting untransformed data in `sjplot` (Lüdtke, 2019)

Predictors	Correct Retrievals			TOTs			Negative TOT			Unsuccessful Retrieval			Unknown Word			Self-Resolved TOTs		
	Log-Odds	CI	p	Log-Odds	CI	p	Log-Odds	CI	p	Log-Odds	CI	p	Log-Odds	CI	p	Log-Odds	CI	p
(Intercept)	1.72	1.30 – 2.14	<0.001	-1.87	-2.20 – -1.54	<0.001	-8.84	-11.54 – -6.13	<0.001	-5.08	-5.89 – -4.26	<0.001	-	-14.63 – -9.44	<0.001	1.00	0.65 – 1.35	<0.001
Gesture inhibited	-0.07	-0.26 – 0.12	0.469	0.02	-0.17 – 0.21	0.834	1.16	-0.48 – 2.81	0.167	0.10	-0.30 – 0.51	0.616	-0.22	-1.06 – 0.63	0.614	-0.03	-0.38 – 0.32	0.869
Random Effects																		
σ^2	3.29			3.29			3.29			3.29			3.29			3.29		
τ_{00}	2.61	Item.Number		1.56	Item.Number		4.19	Item.Number		4.02	Item.Number		37.27	Item.Number		0.34	Item.Number	
	0.48	ID.number		0.25	ID.number		1.46	ID.number		0.32	ID.number		6.11	ID.number		0.28	ID.number	
ICC	0.48			0.36			0.63			0.57			0.93			0.16		
N	34	ID.number		34	ID.number		34	ID.number		34	ID.number		34	ID.number		34	ID.number	
	104	Item.Number		104	Item.Number		104	Item.Number		104	Item.Number		104	Item.Number		93	Item.Number	
Observations	3536			3536			3536			3536			3536			698		
Marginal R ² / Conditional R ²	0.000 / 0.484			0.000 / 0.356			0.036 / 0.645			0.000 / 0.569			0.000 / 0.930			0.000 / 0.157		

Table D. *Fixed and random effects for self-resolved TOTs in both conditions in Experiment 2. Table generated using the tab_model function reporting untransformed data in sjplot (Lüdtke, 2019)*

<i>Predictors</i>	Self-Resolved TOTs in Both Conditions		
	<i>Log-Odds</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.00	0.66 – 1.33	<0.001
Gesture Condition	-0.04	-0.39 – 0.31	0.825
Representational Gesture	0.10	-0.13 – 0.32	0.394
Spatial Span	0.22	-0.01 – 0.46	0.058
Digit Span	0.10	-0.14 – 0.34	0.423
Random Effects			
σ^2	3.29		
τ_{00} Item	0.33		
τ_{00} ID	0.20		
ICC	0.14		
N ID	34		
N Item	93		
Observations	698		
Marginal R ² / Conditional R ²	0.019 / 0.155		

Table E. *Fixed and random effects for self-resolved TOTs in the Gesture Allowed condition in Experiment 2. Table generated using the `tab_model` function reporting untransformed data in `sjplot` (Lüdtke, 2019)*

Self-Resolved TOTs in Gesture Allowed Condition			
<i>Predictors</i>	<i>Log-Odds</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.90	0.63 – 1.17	<0.001
TOT Resolution Gesture Inhibited	0.33	0.07 – 0.59	0.013
Representational Gesture	-0.01	-0.23 – 0.20	0.901
Spatial Span	0.11	-0.13 – 0.35	0.372
Digit Span	-0.20	-0.47 – 0.06	0.134
Random Effects			
σ^2	3.29		
τ_{00} Item	0.11		
ICC	0.03		
N _{Item}	78		
Observations	356		
Marginal R ² / Conditional R ²	0.034 / 0.065		