




# The role of motor context in the beneficial effects of hand gesture on memory

Kimberly M. Halvorson<sup>1</sup> · Alexa Bushinski<sup>1</sup> · Caitlin Hilverman<sup>2</sup> 

Published online: 1 May 2019  
© The Psychonomic Society, Inc. 2019

## Abstract

Viewing co-speech hand gestures with spoken phrases enhances memory for phrases, as compared to when the phrases are presented without gesture. Prior work investigating the mechanism underlying the effect of gesture on memory has implicated engagement of the motor system; when the hands are engaged in an unrelated motor task when viewing gesture, the beneficial effect of gesture is absent. However, one alternative interpretation of these findings is that the beneficial effect of gesture disappears due to mismatched contexts at encoding and retrieval: The hands are engaged during either encoding or retrieval, but not during both stages. Here we examined whether matching the motor context at encoding and retrieval plays a role in the beneficial effect of gesture on memory during a phrase recall task. Participants were presented with phrases that were viewed with and without gesture. Participants were assigned to one of four conditions that determined whether they would complete an unrelated motor task at (1) encoding only, (2) retrieval only, (3) both encoding and retrieval, or (4) neither. During stages in which they were not completing a motor task, participants' hands were in their laps. We found that gesture enhanced memory for phrases both when participants engaged in an unrelated motor task at encoding *and* retrieval and when they did not complete the motor task during either stage. Furthermore, phrases observed with gesture were more likely to be paraphrased than to be recalled literally. Together, these findings demonstrate that gesture can enhance memory even when the motor system is engaged in another task, as long as that same task is performed at retrieval.

**Keywords** Hand gesture · Memory · Motor system · Gesture observation · Perception and action · Embodied cognition

When we communicate, we often produce co-speech hand gestures. Gestures are spontaneous movements of our hands and arms that are semantically related to the content of our speech (McNeill, 1992). These gestures are integrated with the content of spoken language and affect the listener's comprehension of the message (Beattie & Shovelton, 1999; McNeill, Cassell, & McCullough, 1994). Even when the gesture provides information not present in spoken language, listeners remember unique information from the gesture when reporting what they heard (Hilverman, Clough, Duff, &

Cook, 2018; Kelly, Barr, Church, & Lynch, 1999; McNeill et al., 1994). Thus, gesture and spoken language together comprise a unified, dynamic system that can profoundly affect communication and memory.

Observing or producing gesture during encoding can facilitate the learning of a new language in adults (Hilverman, Cook, & Duff, 2018; Kelly, McDevitt, & Esch, 2009; Kroenke, Mueller, Friederici, & Obrig, 2013; Macedonia, 2014). Similarly, in children, observing or producing gesture when learning a new mathematical concept enhances learning for that concept both in an immediate posttest (Cook, Mitchell, & Goldin-Meadow, 2008) and after a delay (Cook, Duffy, & Fenn, 2013). Instructing children to produce gesture when describing a past event enhances their memory for that event (Stevanoni & Salmon, 2005). Therefore, producing or observing gestures during encoding, retrieval, or both can enhance memory.

In the present study, we were specifically interested in the benefit of *observing* gesture on the listener's memory for

✉ Kimberly M. Halvorson  
Kimberly.Halvorson@metrostate.edu

<sup>1</sup> Department of Psychology, Metropolitan State University, St. Paul, MN, USA

<sup>2</sup> Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

spoken words. This benefit derives from a phenomenon known as the *enactment effect* (Cohen, 1989; Engelkamp & Krumnacker, 1980). This effect refers to the finding that pantomiming the relevant movements associated with action words or phrases leads to better memory for those words or phrases. This is true whether the participant is doing the acting or is observing someone else do the acting (Cohen, 1989). Additionally, the enactment effect is found when the enactment occurs during encoding, and a further benefit is found when the same action is performed at retrieval (Engelkamp, Zimmer, Mohr, & Sellen, 1994). Yet the enactment effect only facilitates memory if the action performed matches the verbal content that it is produced with. Action production can *inhibit* memory performance if the action produced does not match the concurrent verbal content. Zimmer and Engelkamp (1984) had participants learn motor action sentences (e.g., “The father is winding up his watch”) and kinematic sentences (e.g., “The smoke was rising”). Participants who produced a concurrent motor action that did not match the content of the sentence (e.g., fist clenching) remembered significantly fewer motor sentences than did participants who had viewed short videos containing kinematic movement (e.g., *a ball rolling across a table*). Thus, unrelated motor engagement while learning sentences inhibited learning, implicating the motor system in the processing of and memory for action sentences, particularly when those sentences also contained motor information.

Similarly, previous studies have implicated the motor system in gesture processing more generally. Spontaneously producing hand gestures while describing a narrative enhances memory for speech as compared to when they are not produced (Cook, Yip, & Goldin-Meadow, 2010). Viewing gesture with sentences facilitates the recall of those sentences, specifically when gestures are related to the verbal information (Feyereisen, 2006). Imaging work has also linked gesture processing and the motor system; a study using EEG demonstrated that having prior sensorimotor experience with an object affects how the gesture for that object is processed (Quandt, Marshall, Shipley, Beilock, & Goldin-Meadow, 2012). Relatedly, a study of both adults and children using fMRI has demonstrated that the neural correlates of observing gesture are affected by how much experience one has in *producing* gesture (Wakefield, James, & James, 2013).

In addition to this empirical work, a theoretical gesture production framework—the gesture as simulated action framework (Hostetter & Alibali, 2008, 2018)—also suggests that the link between gesture and the motor system is critical. According to this account, speakers gesture because they simulate actions and perceptual states while thinking, and these thoughts engage the motor system and serve as the building blocks of gesture. Taken together, this framework and the aforementioned empirical work demonstrate a well-established link between gesture and the motor system. However, the extent to which the motor system is involved

in gesture processing—and specifically in gesture observation’s facilitative effect on memory—remains less clear.

Recent work by Iani and colleagues (Iani & Bucciarelli, 2017, 2018) investigated a direct-activation account of the motor system of the listener as a possible mechanism for the facilitative effect of gesture observation on memory. Similar to Zimmer and Engelkamp (1984), they tested whether having listeners perform a concurrent motor task disrupts the beneficial effect of gesture observation on recall for spoken phrases. Iani and Bucciarelli (2017, 2018) had participants watch videos of a person saying action phrases. In half of the videos, the phrase was accompanied by a meaningful co-speech gesture. For the other half, the phrase was not accompanied by any arm movements. In that case, when there were no additional instructions regarding the participants’ own hands or constraints on his or her movements, recall was better for the action phrases accompanied by gesture (Iani & Bucciarelli, 2017).

In critical comparison conditions, an irrelevant motor task was introduced, such that participants were instructed to move their hands in a rhythmic tapping motion during encoding (i.e., while watching the videos) or during retrieval (i.e., while attempting to recall the action phrases), or they were instructed to move their feet in a comparable pattern at either encoding or retrieval (Iani & Bucciarelli, 2017, 2018). For the conditions in which the hands were engaged in the irrelevant motor task at either encoding or retrieval, the facilitation of memory for action phrases that were accompanied by gesture was disrupted. When listeners’ feet were engaged in an irrelevant task at encoding or retrieval, the benefit for gesture persisted.

Iani and Bucciarelli (2017, 2018) concluded that observing gesture activated the listeners’ own motor systems and that this activation supported the development of a content-rich mental model for representing the speech information. When the hands were engaged in another motor task, the listeners’ motor systems were occupied and unable to encode supplementary information from the gesture. According to this theory, mental models are constructed during discourse, and relevant motor information can contribute to a more fully articulated model. These models can contain both declarative (“knowing that”) and procedural (“knowing how”) knowledge for the to-be-remembered speech (Iani & Bucciarelli, 2017; Iani et al., 2018). Multiple knowledge types foster a more complete understanding of the speech and make it easier to recall.

Despite the central role that the motor system plays in gesture production and observation, manipulating the presence of an additional tapping task at encoding or retrieval results in inconsistencies in the motor context that could also affect recall of the phrases. One recent study has highlighted the importance of the motor context matching in a procedural-learning task that involved instructing participants to gesture or not to gesture. Huff, Maurer, and Merkt (2018) had

participants complete a procedural-learning task (tying knots) and found that participants who gestured during the learning phase were more accurate at test, so long as they *also* gestured during the testing phase. Participants who did *not* gesture during the learning phase were more accurate when they did *not* gesture during the testing phase. Both congruent groups did better than participants who had gestured during learning but not at test (Huff et al., 2018). The researchers suggested that gesturing during learning provides an added benefit for retrieval only when that context is reinstated (i.e., gestures or actual movements are required) at test.

This work is slightly different from the gesture observation studies discussed above, in that participants were not observing someone gesture or making task-irrelevant hand movements, but rather were using their own motor system to mimic the procedural knowledge they were trying to acquire. But the reinstatement of the encoding context at retrieval was crucial to the question we were addressing in the present work. In Iani and Bucciarelli (2017, 2018), manipulating the availability of the listeners' motor systems by engaging the hands and arms with an irrelevant motor task only during encoding or only at retrieval created inconsistencies across the encoding and retrieval contexts. Such inconsistencies are known to be disruptive to memory consolidation and retrieval.

According to the principle of encoding specificity, when a word, for example, is encoded, what is stored is very specific information about that word based on, and including information from, the context from the specific situation in which it was encountered (Tulving & Thomson, 1973). Put more plainly, when an item is encoded into memory, the stored representation is not only the information from the relevant stimulus. Rather, memory includes the information from the stimulus plus any number of situational, environmental, emotional, or semantic cues present at the time of encoding. One example of the reach of the encoding specificity principle is the classic experiment by Baddeley and colleagues in which participants were asked to learn a list of words either on land or under water in full scuba gear (Godden & Baddeley, 1975). In this early demonstration of the role of context on memory formation and retrieval, half of the participants were tested under the same conditions in which they had learned the information, and the other half were tested in the opposite environment. Memory was better when the encoding and retrieval conditions matched—whether that was on land or in water—than when they were different.

In addition to occupying the listeners' motor systems with an irrelevant task, in Iani and Bucciarelli's (2017, 2018) gesture studies the researchers also created different conditions at encoding and retrieval by introducing the motor task only at encoding or only at test. The aim of the present study was to more precisely characterize the conditions under which an irrelevant motor task disrupts the beneficial effect of gesture on recall for action phrases, by keeping the encoding and

retrieval contexts consistent. We hypothesized that providing matching motor contexts at both encoding and retrieval would enhance memory for phrases with gesture, even when the motor system was engaged in a task that was not directly related to the information being learned.

## Present experiment

To address the question of whether the change in context might have had an effect on the benefit of gesture on recall of verbal phrases, we replicated the three primary conditions from Iani and Bucciarelli (2017, 2018), and added a fourth condition in which participants were instructed to perform the irrelevant motor task throughout both encoding and retrieval. In all four conditions, participants saw 24 distinct sentences—12 accompanied by gesture and 12 without. In the first condition—*no tapping*—participants were not told to move their hands in any specific way. In the second condition—*both tapping*—they were instructed to tap at both encoding and retrieval. In the third condition—*encoding tapping*—they were instructed to tap only at encoding. In the fourth condition—*retrieval tapping*—they were instructed to tap only at retrieval. For all four groups, we measured memory for the spoken phrases via an uncued recall task.

We predicted a benefit for sentences accompanied by gesture in two of the four conditions: specifically, the no-tapping and both-tapping conditions. Because the encoding and retrieval contexts were the same in these cases, it was possible that the benefit for memory for sentences accompanied by gesture would be preserved. In the no-tapping condition, following Iani and Bucciarelli (2017), listening to sentences accompanied by gesture might create a more detailed mental model for the phrases. At retrieval, because the motor system was unoccupied, motor simulations for the observed gestures might be reactivated in order to boost the overall number of items remembered.

In the both-tapping condition, we predicted that performing a concurrent motor task would not disrupt the benefit to memory for sentences accompanied by gesture, so long as the same motor context was reinstated at recall. In fact, the continuous activity of the motor system during encoding might allow participants to actually use the tapping task as one of many retrieval cues at recall, since it had been part of the motor context at encoding. By reinstating the encoding context, it was possible that we would see the same benefit for observing gesture on memory for the phrases. Alternatively, if the motor system was overwhelmed by the demands of the tapping task, the motor information from the gestures would never be encoded and therefore would not be available for retrieval, regardless of whether the motor contexts matched.

We made a different prediction for the encoding-tapping and retrieval-tapping conditions. In the encoding-tapping

condition, we predicted no benefit for phrases accompanied by gesture. Assuming that the concurrent motor task was a critical part of the motor context at encoding, information from the gesture-accompanied phrases would be harder to access at recall. For the retrieval-tapping condition, the addition of a new task might make the retrieval process more difficult and potentially diminish the extent to which the information that was acquired via gesture could be used as a retrieval cue or activated. This would replicate the two conditions in Iani and Bucciarelli (2017) that showed no benefit of gesture when the encoding and retrieval conditions were mismatched.

## Method

### Participants

Eighty-three volunteers were recruited for participation in this study. Twenty-three of the participants (19 female, three male, one other) were recruited via an electronic subject pool and participated in exchange for course credit in an introductory psychology course in St. Paul, Minnesota. The average participant age was 32 years old (range 18 to 55): One of the participants identified as Hispanic or Latino; two identified as Asian; seven identified as Black, not of Hispanic origin; 12 identified as White, not of Hispanic origin; and one identified as other. The remaining 60 volunteers (23 female, 35 male, two other) were recruited from the same geographical region via snowball sampling, advertisements on social media platforms, and word of mouth, in exchange for one entry in a drawing for a \$25 gift card to an online retailer. The average participant age was 31 years old (range 18 to 66): One identified as Hispanic or Latino; one identified as American Indian or Alaska Native; two identified as Asian; one identified as Black, not of Hispanic origin; 54 identified as White, not of Hispanic origin; and one identified as other.

### Materials

The 24 normed sentences, consisting of action phrases from Iani and Bucciarelli (2017, Exp. 2), were adapted slightly for comprehension and familiarity (see the Appendix). There were 48 videos total, in which the speaker uttered the phrases with or without accompanying gestures depicting the action information (see Fig. 1 for an example). The action phrases were divided into two sets, with 12 sentences in each set. Two versions of each set were recorded (one with the accompanying gestures, and one without). In total, there were four sets of videos: gesture + Set A, no gesture + Set A, gesture + Set B, and no gesture + Set B. These video sets were used to construct four protocols. In Protocol 1, gesture + Set A was followed by no gesture + Set B. In Protocol 2, gesture + Set

B was followed by no gesture + Set A. In Protocol 3, no gesture + Set A was followed by gesture + Set B. In Protocol 4, no gesture + Set B was followed by gesture + Set A. Thus, the order of the gesture block and the set of action phrases was fully counterbalanced across participants. Stimulus videos were presented on an 13.3-in. monitor using OpenSesame's `media_player_mpy` plugin, which is based on the MoviePy software.

### Procedure

Participants were randomly assigned to one of the four protocols; they were also randomly assigned to one of the four tapping conditions. The same tapping instructions and task from Iani and Bucciarelli (2017) were used here. Participants in the *no-tapping* condition ( $n = 21$ ) were not given any additional instructions about what to do with their hands during either encoding or retrieval. Participants in the *encoding-tapping* condition ( $n = 21$ ) were instructed at the start of each video set to “Place their hands on their knees. Throughout the videos, continuously and alternately tap the table in front of you with your index fingers. After tapping the table with one hand, that same hand would come back down to the knee before the next hand goes on to tap the table.” The 12 sentences in the video set were presented randomly in immediate succession, one after the other. See Fig. 2 for a schematic of the procedure. Next, a white screen appeared with the word “Now” in the center of the screen in black type for 90 s. Participants were asked to recall as many of the phrases as possible (but they were not engaged in the tapping task). Vocal responses were recorded. Participants were then instructed to resume the tapping task while they watched the second video set, which in turn was followed by the “Now” screen. Participants then had 90 s to recall as many sentences as possible. In the *retrieval-tapping* condition ( $n = 21$ ), participants were not told what to do with their hands while watching the videos. They were given the instructions for the tapping task prior to the start of the study and were prompted to begin the tapping task when the “Now” screen appeared. In the *both-tapping* condition ( $n = 20$ ), participants were given the instructions for the tapping task at the start of the experiment and were told to continue the tapping throughout the duration of the study. In total, there were eight possible conditions; the design was fully counterbalanced.

### Coding of the recollections

We adopted the exact same coding system as had Iani and Bucciarelli (2017). Responses were coded to one of three categories: literal recollections, paraphrase recollections, or erroneous recollections. *Literal recollections* were phrases recalled exactly as they had originally been uttered by the speaker. *Paraphrase recollections* were phrases recalled that





**Fig. 1** Stills from videos accompanying the phrase “playing the piano.” The gesture-observed condition is on the left, and the no-gesture-observed condition is on the right. Participants saw two sets of 12 sentences, with one set accompanied by gesture and the other set with the hands at rest

captured the general meaning or gist of the original phrase. We used the same system for identifying paraphrases as had Iani and Bucciarelli (2017), which included changes to the plurality of the items in the phrases, different articles, or minor verb modifications. All other recollections were recorded as *erroneous recollections*. In addition to coding responses, we also identified the phrases that each participant had missed; we incorporated these missed trials into our analyses reported below.

## Results

Participants correctly recalled a mean of 5.70 sentences per set of 12 ( $SD = 1.75$ , range = 2–10 phrases; Fig. 3). The condition with the highest average of correctly recalled phrases was the no-tapping, gesture-observed condition ( $M = 7.00$ ), whereas the condition with the lowest average of correctly recalled phrases was the both-tapping, no-gesture-observed condition ( $M = 5.05$ ).

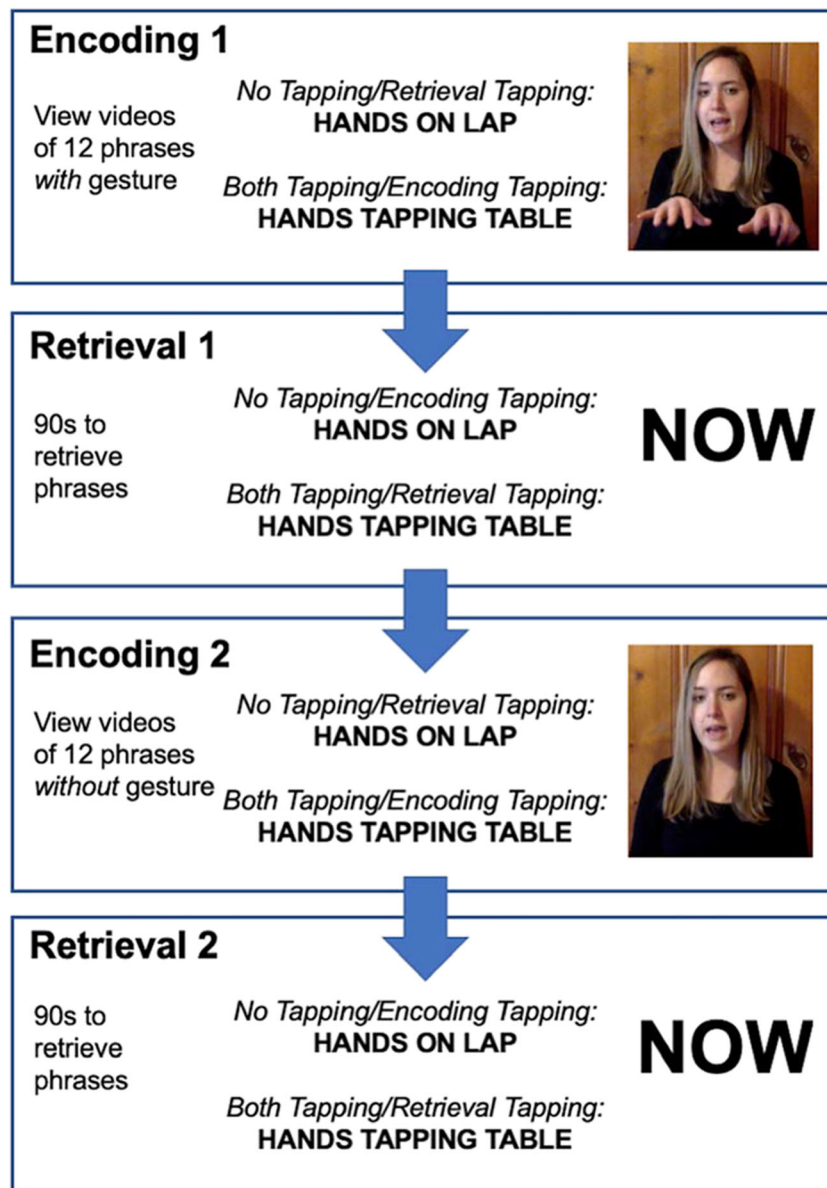
Of those sentences coded as correct, a mean of 4.31 ( $SD = 1.68$ ) were literal recollections, and 1.38 ( $SD = 1.25$ ) were paraphrased. Literal responses comprised 35.2% of all response types, and paraphrased responses comprised 11.4% of response types. Errors were relatively infrequent, with a mean of 0.30 ( $SD = 0.71$ ) per set of 12, comprising just 2.5% of all response types. Missed phrases made up the largest proportion of possible response types, at 50.9%. See Fig. 4 for the breakdown of proportions of responses by condition.

To assess whether context modulated the effect of gesture on memory, we used binomial mixed-effect regression models. We used the `glmer()` function from the `lme4` package (version 1.1-13) in R (version 1.1.419). Tapping condition and gesture type were dummy-coded, with the no-tapping condition and gesture viewed serving as the reference groups. We

determined the random-effect structure by using the most maximal random-effect structure that would converge (Barr, Levy, Scheepers, & Tily, 2013). We ran six models: the first predicting correct recall of a phrase across all four conditions; the second and third predicting correct recall across the two conditions with matching context (no tapping, both tapping), and then the two conditions with mismatching context (encoding tapping, retrieval tapping); the fourth predicting *literal* recall of a phrase across all conditions; the fifth predicting *paraphrased* recall of a phrase across all conditions; and the sixth predicting *errors* in recall of a phrase across all conditions. The model predicting erroneous responses did not converge, due to data sparsity. For the first five logistic regression models, missed and error responses were coded as 0s. For the sixth model, errors were coded as 1s, and the rest as 0s.

## Correct phrase recall

Our first model predicted correct phrase recall—collapsing across literal and paraphrased responses—as a function of tapping condition (no tapping, encoding tapping, retrieval tapping, both tapping), gesture type (gesture or no gesture viewed), and their interactions. There were random intercepts for item and subject, with a random slope for gesture type on the intercept for phrase; the three models we tried with more complex random-effect structure failed to converge. Our final model was  $\text{Recalled} \sim \text{Tapping Condition} * \text{Gesture Type} + (1 + \text{Gesture Type} | \text{Phrase}) + (1 | \text{Subject})$ . We found a main effect of gesture type ( $B = -0.40$ ,  $z = -1.97$ ,  $p = .049$ ); phrases viewed with gesture were more likely to be recalled than those viewed without. There were main effects of tapping condition for both encoding tapping ( $B = -0.57$ ,  $z = -2.86$ ,  $p = .004$ ) and retrieval tapping ( $B = -0.51$ ,  $z = -2.64$ ,  $p = .008$ ); phrases



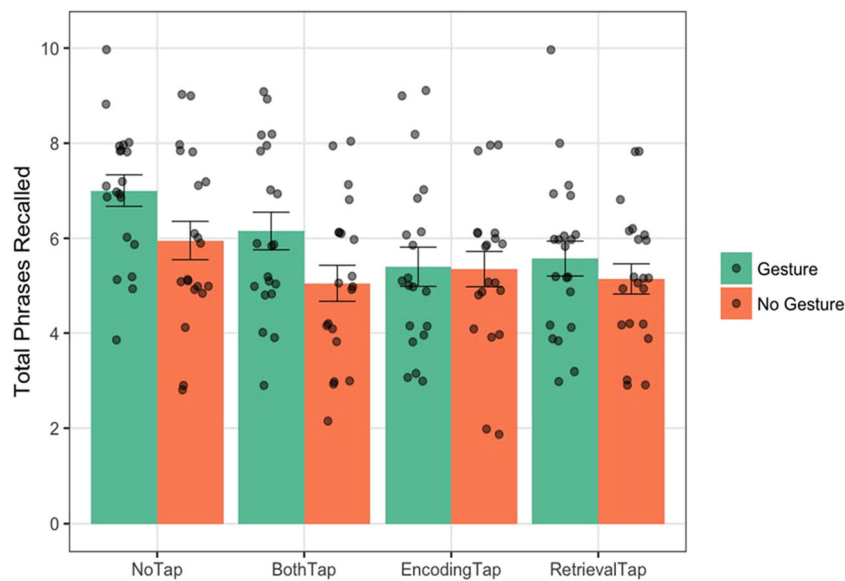
**Fig. 2** Experimental procedure for all conditions. Participants were instructed either to keep their hands on their lap or to rhythmically tap the table in front of them on each phase, depending on the condition to

which they were assigned. The order of the phrase sets presented was randomized across participants; the order of the type of phrase (with gesture or without) was counterbalanced

were more likely to be recalled in the no-tapping condition than in the encoding-tapping and retrieval-tapping conditions. No significant difference emerged between the no-tapping and both-tapping conditions ( $B = -0.31, z = -1.55, p = .12$ ). None of the interactions were significant (Both Tap  $\times$  No Gesture:  $B = 0.03, z = 1.51, p = .88$ ; Encoding Tap  $\times$  No Gesture:  $B = 0.40, z = 1.51, p = .13$ ; Retrieval Tap  $\times$  No Gesture:  $B = 0.26, z = 0.99, p = .32$ ).

After we found no difference between the no-tapping and both-tapping conditions, our next model included just the conditions with matching context. We included this model in order to examine whether these two conditions were different without the added variability present in the full model from the

mismatched conditions. The model structure was the same as that above. We found a marginal main effect of gesture type ( $B = -0.41, z = -1.85, p = .06$ ); phrases viewed with gesture were more likely to be recalled than those viewed without. The main effect of condition was not significant ( $B = -0.31, z = -1.54, p = .12$ ), nor was the interaction ( $B = 0.04, z = 0.16, p = .88$ ). We then ran an identical model with just the encoding-tapping and retrieval-tapping conditions. The main effect of gesture type was not significant ( $B = -0.02, z = -0.53, p = .60$ ). The main effect of tapping condition was also not significant ( $B = 0.05, z = 0.28, p = .78$ ), nor was the interaction of tapping condition and gesture type ( $B = -0.14, z = -0.52, p = .60$ ).



**Fig. 3** Mean numbers of phrases correctly recalled, by tapping condition and gesture type. Participants recalled significantly more phrases in the no-tapping condition than in the encoding-tapping and retrieval-tapping conditions

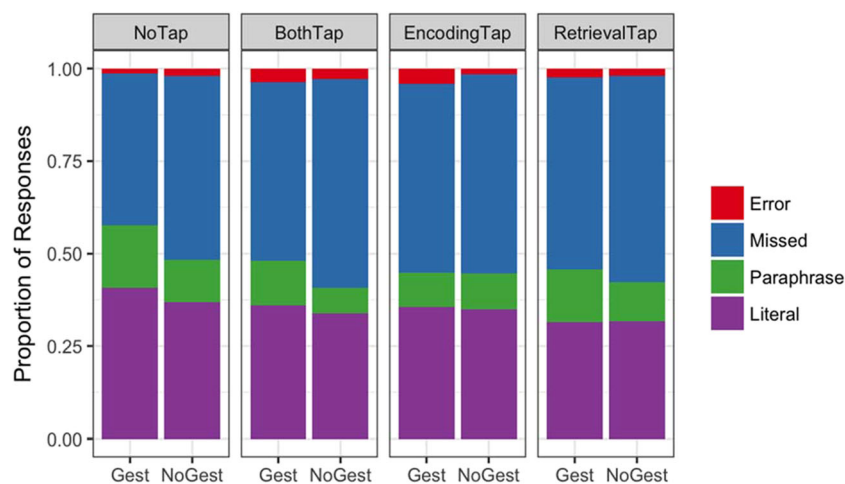
### Literal phrase recall

We next analyzed literal phrase recall as a function of tapping condition and gesture type. The model structure was the same as that above. We observed a main effect of tapping condition ( $B = -0.43$ ,  $z = -2.01$ ,  $p = .045$ ); phrases were more likely to be literally recalled in the no-tapping than in the retrieval-tapping condition. The remaining main effects of both tapping ( $B = -0.11$ ,  $z = -0.53$ ,  $p = .60$ ) and encoding tapping ( $B = -0.25$ ,  $z = -1.19$ ,  $p = .23$ ) were not significant. The main effect of gesture type was not significant ( $B = -0.16$ ,  $z = -0.78$ ,  $p = .43$ ). The interactions of tapping condition and gesture type were also not significant (Both Tapping  $\times$  No Gesture:  $B = 0.005$ ,  $z = 0.27$ ,  $p = .99$ ; Encoding Tapping  $\times$  No Gesture:  $B =$

$0.14$ ,  $z = 0.50$ ,  $p = .62$ ; Retrieval Tapping  $\times$  No Gesture:  $B = 0.17$ ,  $z = 0.60$ ,  $p = .55$ ).

### Paraphrased phrase recall

We next analyzed paraphrased phrase recall as a function of tapping condition and gesture type. The model structure was the same as that above. A main effect of gesture type was apparent ( $B = -0.58$ ,  $z = -2.07$ ,  $p = .038$ ); phrases were more likely to be paraphrased when viewed with gesture than when viewed without. There was also a main effect of tapping condition for the encoding-tapping condition ( $B = -0.78$ ,  $z = -2.38$ ,  $p = .017$ ); phrases were more likely to be paraphrased in the no-tapping than in the encoding-tapping condition. The



**Fig. 4** Proportions of responses for phrases at recall, by tapping condition and gesture type. Participants were significantly more likely to provide a paraphrased response when the phrases were encoded while observing gesture, as compared to those encoded without gesture

remaining effects of tapping condition were not significant (both tapping:  $B = -0.45$ ,  $z = -1.42$ ,  $p = .16$ ; retrieval tapping:  $B = -0.27$ ,  $z = -0.89$ ,  $p = .38$ ). The interactions of tapping condition and gesture type were also not significant (Both Tapping  $\times$  No Gesture:  $B = -0.09$ ,  $z = -0.22$ ,  $p = .83$ ; Encoding Tapping  $\times$  No Gesture:  $B = 0.58$ ,  $z = 1.38$ ,  $p = .17$ ; Retrieval Tapping  $\times$  No Gesture:  $B = 0.21$ ,  $z = 0.52$ ,  $p = .60$ ).

## Errors

We next analyzed errors in recall as a function of tapping condition and error type. We included a random intercept for participant; this was the most complex model that would converge, due to the relatively infrequent occurrence of errors. None of the main effects of tapping condition (both tapping:  $B = 0.958$ ,  $z = 1.19$ ,  $p = .23$ ; encoding tapping:  $B = 1.35$ ,  $t = 1.71$ ,  $p = .09$ ; retrieval tapping:  $B = 0.67$ ,  $t = 0.82$ ,  $p = .41$ ) and phrase type (no gesture:  $B = 0.50$ ,  $t = 0.68$ ,  $p = .50$ ) were significant, nor were their two-way interactions (Both Tapping  $\times$  No Gesture:  $B = -0.77$ ,  $t = -0.85$ ,  $p = .40$ ; Encoding Tapping  $\times$  No Gesture:  $B = -1.46$ ,  $t = -1.52$ ,  $p = .13$ ; Retrieval Tapping  $\times$  No Gesture:  $B = -0.72$ ,  $t = -0.74$ ,  $p = .46$ ).

## Discussion

We investigated whether viewing gesture enhances memory for phrases and whether engaging in an unrelated motor task mitigates the effect of gesture on memory, depending on whether the encoding and retrieval contexts match. Consistent with prior work, we found that seeing and hearing phrases accompanied by gesture enhanced memory for those phrases (Cohen, 1989; Engelkamp et al., 1994; Iani & Bucciarelli, 2017, 2018). Furthermore, we replicated Iani and Bucciarelli (2017, 2018) by demonstrating this beneficial effect of gesture in the no-tapping condition, but we did not observe it in the encoding-tapping or retrieval-tapping conditions. As predicted, the results of the new, both-tapping condition, which had not been included in the previous studies by Iani and Bucciarelli (2017, 2018), showed that the motor contexts at encoding and retrieval mattered: Performance was affected by whether the encoding and retrieval contexts matched. Specifically, participants who engaged in a motor task at encoding *and* retrieval performed similarly to participants who did not engage in a motor task at either stage. Therefore, engaging in an unrelated motor task does not wipe out any beneficial effect of gesture for memory. Rather, matching the learning and recall contexts—by engaging the hands and arms in the same task at encoding and retrieval—leads to enhanced memory for phrases accompanied by gesture, as compared to those unaccompanied by gesture.

Although our first model comparing all four conditions did not yield significant interactions, the main effects of the encoding and retrieval conditions suggested different performance based on whether or not the contexts matched. We ran a follow-up model only on the no-tapping and both-tapping conditions, to directly compare the novel condition with the baseline condition. We found that gesture significantly enhanced memory for phrases in both conditions, but the overall mean phrases recalled were not significantly different across groups. When we ran this same model in the encoding-tapping and retrieval-tapping conditions, the beneficial effect of gesture was not present. We can conclude from these results that even when the motor system is engaged in a motor task during encoding, information is still encoded from gesture. If the motor system is engaged in that same task at retrieval, a benefit for gestured information persists.

How is it that engaging in a motor task at both encoding and retrieval showed the same facilitative effect of gesture on recall as keeping the hands at rest? We do not interpret these findings as evidence that the listeners' motor system is not involved in gesture observation, understanding, or comprehension. Rather, we suggest that engaging the arms and hands in a secondary motor task during encoding or retrieval does not, on its own, disrupt the benefit of gesture. Although the tapping task was unrelated to the primary task of recalling the spoken phrases, we interpret our findings as evidence that the tapping was never task-irrelevant. It seems that instructing participants to engage in the tapping task created a motor context that must be present at both encoding and retrieval for the benefit of gesture to be observed. We argue that unlike the critical control condition from Iani and Bucciarelli (2018), in which participants were prompted to move their legs and feet in a secondary motor task, moving your hands while watching someone else move their hands is inherently task-relevant. The reason that moving your feet does not disrupt the beneficial effect of gesture is because it truly is task-irrelevant.

We can conceive of a few explanations for the persistence of the benefit for gesture in the both-tapping condition, based on potential differences in the roles of the listeners' motor systems during gesture observation once the relevant motor context from encoding has been reinstated at retrieval. According to one view, gesture observation elicits an identical motor trace in the motor system of the listener; in this case, we think it is entirely possible that the motor information from the gestures was simply integrated with the ongoing motor activity from the tapping task. Given that the tapping task used was relatively simple, rhythmic, and repetitive, it likely was not a significant strain on the cognitive and neural systems underlying action planning and production. It could be that plenty of resources were available in working memory, such that the information from gesture could be stored along with or in addition to the action information necessary for executing the tapping task.



When the same movements were produced again at recall, this could cue the action information encoded concurrently via gesture. This reactivation of the motor information might serve as an effective retrieval cue that made the verbal information easier to access or the memory for the spoken information more robust. In this case, the action information and verbal information from each phrase could be stored (and subsequently reactivated) as a single, multimodal representation.

It is also possible that during gesture observation, the listener acquires *additional* content from the gestures via a motor simulation, but ultimately this information does not get stored as a motor trace. Instead, motor simulation during gesture observation facilitates comprehension of the spoken content by activating or even generating a corresponding mental image or other analog representation of the to-be-remembered information. This is consistent with the gesture as simulated action framework (Hostetter & Alibali, 2008, 2018) described earlier.

We remain agnostic as to the precise role of the listener's motor system in gesture observation. However, as we noted above, the tapping task used in these experiments was repetitive and simple, arguably not putting great burden on the motor system. Would engaging the hands in a more complicated motor task have overloaded the motor system and wiped out the benefit of gesture? There is some evidence that this would be the case. Ping, Goldin-Meadow, and Beilock (2013) had participants view and hear sentences—some accompanied with gesture and some not—and then subsequently make judgments on whether a pictured object had been present in that sentence. Gestures contained additional information that could be used to speed up reaction times for the judgments. One group of participants completed this task while carrying out a concurrent motor task involving making unplanned hand and arm movements. Ping et al. found that engaging in this motor task rendered participants incapable of using the information from gesture to influence their sentence comprehension. It remains unknown whether engaging participants in a more complex motor task would mitigate the effect of gesture on memory. Follow-up studies could engage participants in a more complex task to investigate whether the benefit for gesture would be eliminated in a both-tapping-like condition that utilized a complex motor task.

Another aspect of our findings that speaks to mechanism is the relative proportions of paraphrased versus literal recollections of the phrases by gesture type; gesture's facilitative effect on memory appears to have been driven by the paraphrased responses. When we restricted the analysis just to paraphrased recollections, we found a significant effect of gesture on memory. When we restricted the analysis just to literal recollections, this benefit disappeared. This is consistent with earlier findings showing a detriment in memory when cued with literal phrases at recall. Cutica and colleagues had both

children (Cutica, Iani, & Bucciarelli, 2014) and adults (Cutica & Bucciarelli, 2013) read scientific texts, and then gave them tests measuring comprehension, verbatim memory, and inference-based questions. They found that when participants were instructed to gesture during encoding, they got more questions correct at test than when they did not gesture. However, in a follow-up experiment, the participants who had been instructed to gesture got *fewer* questions correct on a recognition test when they included a literal phrase from the study materials than when they had not been instructed to gesture (Cutica & Bucciarelli, 2013). The researchers concluded that the beneficial effect of gesture improves memory by helping establish a more detailed, articulated mental model for the written information. One side effect of this process may be a diminished memory for surface features of the original material (Cutica & Bucciarelli, 2013).

We posit a similar interpretation of our results: Observing gesture with the phrases significantly enhanced memory for paraphrased, but not literal, responses. We suggest that this may have been due to participants sometimes relying more heavily on their memory for the gesture than on the phrase itself. For example, the phrase “whisking eggs” was presented with a gesture of a hand moving vigorously in a circular motion in the gesture condition. At recall, the reactivation of the motor information from gesture could link back to several different ways to phrase this in spoken language (e.g., *whisking some eggs*, *beating eggs*, *stirring the eggs*, etc.). Because specific gestures can map on to multiple different words and ways of phrasing the intended meaning or can activate corresponding images or analog representations of the semantic content, retrieving multimodal representations via gesture is likely to lead to a “gist” memory for what was encoded from spoken language. Indeed, the phrases that had the lowest incidence of paraphrased responses were all phrases that have clear mappings with specific gestures; the gesture canonically represented the literal phrase, with few other options for what it was representing (i.e., *hammering a nail*, *rowing a boat*, or *shooting a gun*). Future work should further investigate this possibility by using gesture and phrase pairings that vary systematically with respect to how clearly the gesture represents the literal phrase. Understanding this distinction has practical implications for the use of gesture in classrooms, therapeutic environments, and other learning contexts; this may suggest that viewing gesture is most useful when the goal is to learn and understand concepts that do not require rote memory for specific words.

In sum, we followed up on prior work demonstrating that a concurrent motor task diminished the effect of gesture on memory for phrases, by testing a critical condition: matching the learning and retrieval contexts by engaging in a motor task at both encoding and recall. We found that participants who completed a tapping task with their hands at both stages performed similarly to participants who did not engage in the task

at either stage. Furthermore, we found that when participants viewed gesture with phrases, they were more likely to provide paraphrased responses than when they did not view gesture. Taken together, these findings suggest that the learning context—specifically, the motor context for the task-relevant effectors—is critical for assessing how gesture affects memory for phrases. Engaging in an unrelated motor task need not disrupt the benefit to memory of observing gestures if the task is completed at both encoding and retrieval.

## Appendix: Stimulus phrases

### Rowing a boat

Conducting an orchestra  
 Playing the violin  
 Dribbling a basketball  
 Playing the piano  
 Cleaning a window  
 Driving the car  
 Painting a painting  
 Ironing a shirt  
 Whisking eggs  
 Wringing out a washcloth  
 Throwing a stone  
 Getting shampoo  
 Polishing silver  
 Hammering a nail  
 Brushing your teeth  
 Putting lotion on your hands  
 Stacking some blocks  
 Sewing by hand  
 Typing on a computer  
 Giving a hug  
 Shooting a gun  
 Rolling some yarn  
 Sharpening a knife

## References

- 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*, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology*, *18*, 438–462. <https://doi.org/10.1177/0261927X99018004005>
- Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review*, *1*, 57–80. <https://doi.org/10.1007/BF01326550>
- Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, *84*, 1863–1871. <https://doi.org/10.1111/cdev.12097>
- Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*, 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*, 465–475. <https://doi.org/10.1016/j.jml.2010.07.002>
- Cutica, I., & Bucciarelli, M. (2013). Cognitive change in learning from text: Gesturing enhances the construction of the text mental model. *Journal of Cognitive Psychology*, *25*, 201–209. <https://doi.org/10.1080/20445911.2012.743987>
- Cutica, I., Iani, F., & Bucciarelli, M. (2014). Learning from text benefits from enactment. *Memory & Cognition*, *42*, 1026–1037. <https://doi.org/10.3758/s13421-014-0417-y>
- Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *27*, 511–533.
- Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition*, *22*, 34–39. <https://doi.org/10.3758/BF03202759>
- Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, *18*, 185–205. <https://doi.org/10.1080/09541440540000158>
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, *66*, 325–331.
- Hilverman, C., Clough, S. A., Duff, M. C., & Cook, S. W. (2018). Patients with hippocampal amnesia successfully integrate gesture and speech. *Neuropsychologia*, *117*, 332–338. <https://doi.org/10.1016/j.neuropsychologia.2018.06.012>
- Hilverman, C., Cook, S. W., & Duff, M. C. (2018). Hand gestures support word learning in patients with hippocampal amnesia. *Hippocampus*, *28*, 406–415. <https://doi.org/10.1002/hipo.22840>
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*, 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Hostetter, A. B., & Alibali, M. W. (2018). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, 1–32. <https://doi.org/10.3758/s13423-018-1548-0>
- Huff, M., Maurer, A. E., & Merkt, M. (2018). Producing gestures establishes a motor context for procedural learning tasks. *Learning and Instruction*, *58*, 245–254. <https://doi.org/10.1016/j.learninstruc.2018.07.008>
- Iani, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110–121. <https://doi.org/10.1016/j.jml.2017.05.004>
- Iani, F., & Bucciarelli, M. (2018). Relevance of the listener's motor system in recalling phrases enacted by the speaker. *Memory*, *26*, 1084–1092. <https://doi.org/10.1080/09658211.2018.1433214>
- Iani, F., Burin, D., Salatino, A., Pia, L., Ricci, R., & Bucciarelli, M. (2018). The beneficial effect of a speaker's gestures on the listener's memory for action phrases: The pivotal role of the listener's premotor cortex. *Brain and Language*, *180–182*, 8–13. <https://doi.org/10.1016/j.bandl.2018.03.001>
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*, 577–592. <https://doi.org/10.1006/jmla.1999.2634>
- Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language & Cognitive Processes*, *24*, 313–334. <https://doi.org/10.1080/01690960802365567>
- Kroenke, K.-M., Mueller, K., Friederici, A. D., & Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex*, *49*, 2553–2568. <https://doi.org/10.1016/j.cortex.2012.11.016>

- Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*, 1467:1–6. <https://doi.org/10.3389/fpsyg.2014.01467>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, *27*, 223–237. [https://doi.org/10.1207/s15327973rlsi2703\\_4](https://doi.org/10.1207/s15327973rlsi2703_4)
- Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2013). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, *143*, 195–204. <https://doi.org/10.1037/a0032246.Understanding>
- Quandt, L. C., Marshall, P. J., Shipley, T. F., Beilock, S. L., & Goldin-Meadow, S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: An EEG study of action production and gesture observation. *Neuropsychologia*, *50*, 2745–2751. <https://doi.org/10.1016/j.neuropsychologia.2012.08.005>
- Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior*, *29*, 217–233. <https://doi.org/10.1007/s10919-005-7721-y>
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. <https://doi.org/10.1037/h0020071>
- Wakefield, E. M., James, T. W., & James, K. H. (2013). Neural correlates of gesture processing across human development. *Cognitive Neuropsychology*, *30*, 58–76. <https://doi.org/10.1080/02643294.2013.794777>
- Zimmer, H. D., & Engelkamp, J. (1984). Planungs- und Ausführungsanteile motorischer Gedächtniskomponenten und ihre Wirkung auf das Behalten ihrer verbalen Bezeichnungen [Planning and execution elements of motor memory components and their effect on the retention of their verbal descriptions]. *Zeitschrift für Psychologie mit Zeitschrift für Angewandte Psychologie*, *192*, 379–402.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.