ARTICLE

Comparing the cognitive load of gesture and action production: a dual-task study

Autumn B. Hostetter and Sonal Bahl

Department of Psychology, Kalamazoo College, Kalamazoo, MI, USA
Corresponding author: Autumn B. Hostetter; Email: autumn.hostetter@kzoo.edu

Abstract

Speech-accompanying gestures have been shown to reduce cognitive load on a secondary task compared to speaking without gestures. In the current study, we investigate whether this benefit of speech-accompanying gestures is shared by speech-accompanying actions (i.e., movements that leave a lasting trace in the physical world). In two experiments, participants attempted to retain verbal and spatial information from a grid while describing a pattern while gesturing, while making the pattern, or while keeping hands still. Producing gestures reduced verbal load compared to keeping hands still when the pattern being described was visually present (Experiment 1), and this benefit was not shared by making the pattern. However, when the pattern being described was not visually present (Experiment 2), making the pattern benefited verbal load compared to keeping hands still. Neither experiment revealed a significant difference between gesture and action. Taken together, the findings suggest that moving the hands in meaningful ways can benefit verbal load.

1. Introduction

Speakers often produce hand movements that depict the meaning of their speech. Such movements, termed iconic gestures (or hereafter, simply gestures), can benefit listeners' understanding of the speaker's message (e.g., Hostetter, 2011) and can also have positive cognitive effects for the speakers who produce them (Goldin-Meadow & Alibali, 2013). Gestures have been shown to facilitate conceptual planning of a message (e.g., Kita & Davies, 2009), assist with lexical access to spoken words and phrases (e.g., Krauss, 1998), and relieve working memory demands (e.g., Goldin-Meadow et al., 2001) for the speakers who produce them. However, while much work has examined the cognitive consequences associated with speaking with gesture compared to those associated with speaking without gesture (e.g., Ping & Goldin-Meadow, 2010), few studies have compared the cognitive effects of producing speech-accompanying gesture to producing speech-accompanying action. The purpose of

the present study is to examine the consequences of gesture versus action on a speaker's cognitive load.

Cognitive load refers to the amount of burden placed on the working memory system by a task. Following Baddeley and Hitch (1974), working memory is conceptualized as a limited resource that can store and manipulate a finite amount of information at any one time. Although individuals differ in how much information their working memory system can manage (e.g., Barrett et al., 2004; Jarrold & Towse, 2006), there are also task differences, with some tasks imposing more cognitive load than others (e.g., Turner & Engle, 1989). Cognitive load is often measured by using a dual-task paradigm, in which participants are asked to complete a primary task under the load of a secondary task. Performance on the secondary task is taken as a measure of the load imposed by the primary task; if participants do well on the secondary task, it implies that the load of the primary task was low and they had many resources available to devote to the secondary task.

Using this type of dual-task paradigm, gestures have been shown to reduce the cognitive load involved in speaking compared to speaking about the same content without gestures. Goldin-Meadow et al. (2001) asked participants to describe their solution to a math problem either with gesture or without (the primary task) while also holding two-letter sequences in memory (the secondary task). They found that participants remembered significantly more of the sequences when they gestured as they described the math problems than when they did not gesture, suggesting that gestures reduced the cognitive load involved in describing the math problems. The effect has been replicated several times (e.g., Cook et al., 2012; Marstaller & Burianova, 2013; Ping & Goldin-Meadow, 2010; Wagner et al., 2004), and the general explanation for the finding is that producing a gesture that represents content of what is being said makes it easier to think about and describe that content, thereby reducing speakers' cognitive load and allowing more cognitive resources to be devoted to a second, unrelated task. Kita et al. (2017)) describe this as a 'downstream' effect of gesture (p. 15); that is, by making it easier to think and speak about the representation, the gesture frees cognitive resources that can be used for another purpose.

Gestures are not the only kind of speech-accompanying behavior that could make thinking and speaking about representations easier. In some situations, speakers can also perform physical actions that enact or depict the content of what is being said. In contrast to a gesture, these speech-accompanying actions change the physical state of the world and leave a lasting, visual impression of the representation. For example, a speaker who gestures about a triangle by tracing the outline of the shape in the air with their index finger leaves no lasting trace of the triangle that can be visually referred to. In contrast, a speaker who makes the triangle they are describing (out of blocks or dough) leaves a visual depiction of the triangle behind that can be consulted to calculate the exact size or angle of the sides, for example. As such, it is possible that actions are advantageous over gesture in some circumstances because actions make the spatial information being described concrete and visible.

Indeed, there is some evidence that actions can have this effect for observers. Listeners perceive actions and gestures differently, with actions resulting in more concrete, less representational interpretations than gestures (Novack et al., 2016). Hostetter et al. (2020) found that listeners who saw a speaker move objects had better memory for where those objects ended up than listeners who saw a speaker gesture about how the objects were moved. Similarly, Kelly et al. (2015) found that listeners

were faster to verify the relation of a target word to an accompanying action than to an accompanying gesture, suggesting that actions were easier for listeners to process than gestures. Actions provide a richer visual code than gesture and seem to provide more cognitive support for observers as a result. Perhaps speakers who produce actions (vs. gestures) also benefit from the more detailed information they provide.

Although such a hypothesis makes some intuitive sense, there is research suggesting that actions are actually *less* beneficial to speakers than gestures in situations where the goal is to generalize beyond the learning context that is being enacted or gestured. Novack et al. (2014) compared the effects of producing a gesture that mimicked the solution to an equivalence problem to the effects of physically moving manipulatives to solve the problem on children's ability to learn the concept of mathematical equivalence. They found that producing gestures led to more success generalizing what had been learned to new situations than producing physical actions – suggesting that gestures are superior to actions because they encourage deeper, more abstract thinking about the most relevant components of an action than actually producing the action. When a speaker produces an action, they must orient their hand to physically grasp the manipulatives and successfully move them; this level of motor control is not necessary when a gesture is produced, leaving the gesturer more opportunity to focus on the meaning of the gesture and how it relates to their speech.

This idea that gestures may be particularly adept at capturing the most relevant pieces of action has been described by Kita et al. (2017), who argue that gestures are a way of *schematizing* action. Under their view, gestures enact the most relevant, schematic parts of an action that are meaningful in a particular context (e.g., the spatial configuration of the triangle's sides) and may drop the other pieces of the action that are not relevant to that context (e.g., hand grasp and hand orientation). Because of this schematizing, gestures effectively highlight the most important components of an action for both speakers and listeners (see also Goldin-Meadow & Beilock, 2010; Novack & Goldin-Meadow, 2017) and enhance performance on a subsequent generalization task where those highlighted components are needed.

Although research supports the idea that gestures may be superior to action for generalization to new contexts, there is less evidence about whether gestures may also enhance learning of the specific information more than action. In one of the few studies to examine this, So et al. (2014) found that gestures were better than drawing for learning spatial routes. Learners either rehearsed spatial routes by tracing them in the air with their finger (Gesture condition), by drawing them on paper with a pencil (Drawing condition), or by thinking about the route without any overt movement (Mental Simulation condition). All three rehearsal conditions led to better recall of the routes than when no rehearsal was allowed, but performance in the Gesture condition was significantly better than performance in the others. It appears that gestures (even more than drawing) may help speakers learn the details of the information they are gesturing about.

One reason why gestures may have this effect on memory for information is because they require more mental effort to produce than action – that is, rather than using cues in the environment (the manipulatives and the scene) to recreate information through action, gestures require that this information be maintained mentally. A speaker who gestures about a spatial scene must keep track of where they are in their description, whereas a speaker who creates the scene through action can refer to the physical layout they are creating to index where they are in the description and

what should come next. Thus, similar to a 'desirable difficulty' that has been described in the memory literature (Bjork & Bjork, 2011), gestures may benefit long-term maintenance and understanding of the information even while costing some resources at the time of production.

Alternatively, it is also possible that gestures require *less* cognitive effort to produce than action. Because gestures do not require the same amount of motor control or precision as actually producing an action that manipulates objects, gestures may allow speakers to express only the elements of their idea that are salient or important in a particular context. Expressing fewer elements in gesture may require less cognitive effort than producing an action that must include even irrelevant information such as hand grasp, hand orientation, and lifting velocity. It is also possible that gestures are easier to produce than actions because speakers simply have more experience in producing gestures alongside speech than actions, and/or because gestures share a unique evolutionary history with speech that is not shared by action. In line with any of these possibilities, there is some evidence that gestures are timed more synchronously with speech than actions (Church et al., 2014), suggesting that gestures may be cognitively easier to plan, produce, and execute than actions.

The present study will examine the cognitive effort involved in producing gestures versus actions in a dual-task paradigm. In line with previous work (e.g., Goldin-Meadow et al., 2001), we predict that gesturing will lead to better performance on a secondary task than speaking with hands still. Further, we examine whether speech-accompanying actions have a similar benefit. If gestures have their beneficial effects on cognition because they schematize only the most important components of a representation, then their benefit to cognitive load should not be shared by action. On the other hand, if gestures have their beneficial effect because they make an internal representation momentarily visible and therefore easier to think and speak about, then their benefit should be shared (and possibly even exceeded) by action.

To test these possibilities, participants in two experiments engaged in a dual-task paradigm that involved describing spatial arrays either with gesture, with action, or with hands still. In Experiment 1, participants described the pattern as it remained visible on the screen. In Experiment 2, participants described the pattern from memory after the pattern disappeared from the screen. In both experiments, participants completed this primary task while also attempting to store both verbal (identity of letters) and spatial (the location of those letters in a grid) information on a secondary task. We measured participants' ability to recall the letters and their ability to recall the locations on each trial as a measure of their verbal and spatial load during the description task.

2. Experiment 1

In a within-subjects design, participants completed 24 trials in which they described a spatial pattern while attempting to maintain verbal and spatial information on a secondary task. They described eight patterns with gesture, eight while making the pattern, and eight while keeping their hands still. We predicted that participants would have the most cognitive capacity available to devote to the secondary task when they gestured with their description, and that this benefit to cognitive load would result in higher recall for both spatial and verbal information on the secondary task on gesture trials. We predicted that action trials (trials on which participants

made the pattern they were describing) might also result in lower cognitive load (i.e., higher recall on the secondary task) than hands-still trials because meaningful actions may also allow speakers to externally visualize what they are describing, making the information easier to think and speak about, in a similar way as gestures.

We also considered aspects of participants' speech during their descriptions (i.e., number of words spoken, number of filled pauses, total amount of time spent speaking, and deictic reference) as control measures. If speakers produce more words and spend more time on some trials than others, then their ability to remember the secondary information from the grid will likely be worse on those trials as well, as it has been longer and there has been more interfering information (e.g., words) since they saw the grid. Further, it is possible that gestures (and possibly actions) have their effect because they reduce the difficulty involved in speech planning and production (e.g., Krauss, 1998), leading to fewer filled pauses on gesture (and action) trials than hands-still trials. This could also be evident in speakers' use of deictic references (words or phrases that refer to the physical environment such as 'like this', 'over here', or 'that one'). Rather than needing to verbalize all of the details, deictic references that refer to gesture, action, or the image on the screen allow the speaker to use simpler language, thereby letting the gesture, action, or image carry more of the load. Use of such references may be more likely in some conditions than others and may also result in increased memory on the secondary task (as the primary description task has been simplified). Thus, we will consider both how manual behavior is related to each speech measure and whether each speech measure predicts memory for the letters or locations. In analyses of how manual behavior predicts memory, we will use a model comparison approach that considers the inclusion of all possible speech variables as covariates.

2.1. Method

2.1.1. Participants

We recruited participants from introductory psychology courses during a single academic term and accommodated all students who wanted to participate. Sixty participants completed this study in exchange for extra credit in their course, but seven were international students who had not learned English from before 5 years of age, and their data were excluded from analysis. A technical problem with the recording equipment led to the loss of data from four more participants. The analyzed sample came from 49 participants (24 men; 25 women) with an average age of 18.85 years ($SD = 0.65$). The majority of participants (57%) identified as white, with Asian (14%), Black (10%), Latinx (10%), and Multiracial (8%) students also represented.

2.1.2. Stimuli

Dot patterns. The primary task was to describe patterns of dots connected with lines to construct geometric shapes and figures. Three sample patterns are shown in Fig. 1. Each pattern contained six or seven black dots with lines drawn to indicate one or two shapes. We adopted the patterns used by Hostetter et al. (2007) and also constructed some new similar patterns to total 27 unique patterns. Although some of the patterns had the same underlying arrangement of dots, all had lines to indicate a unique geometric configuration that needed to be described.
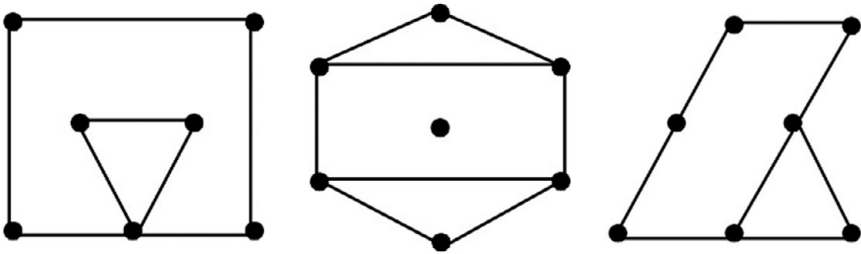
Figure 1. Three sample patterns used in the two experiments.

Grids. The secondary task was to remember the configuration of five letters presented in five locations within a 4 × 4 grid. We constructed 27 such grids for use in the trials, instructions, and practice of the experiment. We attempted to avoid letter combinations that spelled any obvious words or common abbreviations.

2.1.3. Procedure

Participants arrived individually to the lab and were told that the study was about how people communicate and remember information. The lab was equipped with an iMac running PsyScope software (Cohen et al., 1993). After signing the consent form, they were given instructions and two practice trials before beginning the 24 experimental trials.

Each trial began with the presentation of one of the 4 × 4 grids centered on the computer screen for 5 s with a prompt above it that said "Remember this!" Once the grid disappeared, it was replaced by one of the dot patterns centered on the screen. Above the pattern, one of three types of instructions was given (GESTURE, DO NOT GESTURE, or MAKE), depending on the condition of the trial. In gesture trials, participants described the pattern they saw on the screen while producing speech-accompanying gestures that depicted the dots and shapes they were describing. During the instructions, the experimenter modeled two such gestures for participants. Participants completed one practice trial in the Gesture condition, and the experimenter gave feedback if needed. In do not gesture trials, participants described the patterns while keeping their hands as still as possible. In the make trials, participants were provided with seven round wooden pieces with ½-inch diameter. Participants arranged these pieces in the locations of the dots in the pattern as they were describing. During the instructions, the experimenter modeled how to do this. Participants completed one practice trial in the Make condition, and the experimenter gave feedback if needed that emphasized the need to position the pieces and describe the pattern simultaneously. The wooden pieces remained on the table throughout the experimental session in all conditions.

When participants were done with their description, they pressed the spacebar on the computer keyboard. The pattern was then replaced on the screen with a prompt to recall the information from the grid. Participants were provided with an answer page that included 27 blank grids that were numbered. They were instructed to write the five letters in the five locations they had seen before their description in the appropriate grid for that trial. When participants had finished recalling the grid information, they pressed the spacebar to advance to the next trial.

After receiving the instructions and feedback on the two practice trials, participants completed the 24 experimental trials. Grids and dot patterns were presented in the same fixed order to all participants across the 24 trials. Three orders were created that counterbalanced which of the three conditions was instructed on each trial. All orders presented the make instructions for eight trials, the gesture instructions for eight trials, and the no-gesture instructions for eight trials in a pseudo-random non-blocked order. In this way, when participants saw the grid for each trial, they had no way of knowing which of the three instruction types would be presented on the screen with the subsequent pattern. Participants were randomly assigned to one of the three orders, and across all participants, each trial (with its associated grid and dot pattern) was completed approximately equally often in each of the three conditions.

The experimenter remained in the room for the entirety of the session. Following the completion of the final trial, the experimenter debriefed the participant and collected demographic information.

2.1.4. *Data coding, exclusions, and analyses*

The grid for each trial was scored for the number of letters and number of locations filled in correctly. The number of letters was determined by comparing the letters written by the participant against those shown in the grid at the beginning of the trial, regardless of the specific locations the letters were placed in. Scores could range from 0 to 5. Similarly, the number of locations was determined by comparing the locations indicated by the participant against those shown in the grid at the beginning of the trial, regardless of the specific letters that were written in each. Scores could range from 0 to 5. Note that participants' scores for letters and locations on each trial were independent of one another, as participants could write all five correct letters in the wrong locations or five incorrect letters in the correct locations.

We also transcribed and coded participants' behavior during the description of the dot pattern on each trial. We transcribed each description verbatim and calculated the total number of words produced as well as the total amount of time spent speaking from when the first word was spoken until the key was pressed to advance to the grid recall phase of the trial. We also counted the number of filled pauses produced during each description (e.g., um, uh, and er). Last, we coded the speech that occurred on each trial for whether it contained a deictic reference in which the speaker referred to something in the physical environment, whether it be something they were showing in their gesture, an aspect of the shape they were making, or a reference to the shape on the screen. For example, one speaker said "there were two triangles here and here…" as they placed the pieces on the table to make the triangles they were referring to. This description was coded as containing dietic reference. In contrast, another speaker described the same pattern by saying "you have a slanted triangle that's slanted with a little triangle at the back of it…" This description was coded as containing no deictic reference.

Finally, we coded participants' hand and arm movements during each description for speech-accompanying iconic gestures. An iconic gesture was defined as a movement that did not touch or manipulate any physical object (the circular wooden blocks, the participants' clothing, and the computer screen) and that indicated or depicted some aspect of what the participant was describing, either by moving the hands in space, forming a shape with the hands, or pointing to a location in space. Although we also noted beat gestures (movements that seemed to emphasize rhythm

or prosody without demonstrating any obvious iconic meaning) when they occurred, these were not very common in our dataset and were not of interest.

We discarded trials on which participants did not follow the instructions for that trial. We excluded 15 trials in the Gesture condition, seven because the participant did not produce any iconic gestures, four because they made the pattern instead of gesturing about it, two because the experimenter interrupted to clarify the instructions, one because they put six letters in the grid, and one because the trial was inadvertently skipped by a double tap of the space bar. In the no-gesture condition, 31 trials were excluded, 12 because the participant produced one or more iconic gestures, 9 because the participant made all or part of the pattern, 4 because the experimenter interrupted to remind about the instructions, 4 because the participants' hands were out of view of the camera preventing an accurate assessment of whether they were gesturing, and 2 because the trial was inadvertently skipped. We excluded 79 trials in the Make condition. The majority of these were because the participant gestured at least one time in addition to making the pattern ($n = 23$) or because the participant did not synchronize the actions involved in making the pattern with their speech ($n = 50$). Specifically, participants in some instances made the pattern completely and then described it, or they continued talking after the pattern had been made. In addition, two trials in the Make condition were excluded because the participant treated it as a no-gesture trial, two because the experimenter interrupted with a reminder about the instructions, one because it was inadvertently skipped, and one because the video stopped recording in the middle of the trial. In total, these exclusions led to the loss of 10.6% of the data. The final dataset included an average of 7.67 trials per participant in the Gesture condition (Range: 6–8), 6.50 trials per participant in the Make condition (Range: 3–8), and 7.37 trials per participant in the Hand Still condition (Range: 5–8).

To calculate reliability for determining whether an iconic gesture occurred on a trial or not, a second coder examined the manual behavior on a random selection of 5% of the trials. This coder was blind to the condition of the trial (e.g., what instructions were on the screen) and simply judged whether the participant produced an iconic gesture on the trial or not. Agreement between the two coders for whether a gesture occurred was substantial (90% agreement; Cohen's $k = .796$). Decisions about whether to exclude a trial were made on the basis of the first coder's judgments.

For analysis, we first considered whether the type of trial (gesture, no gesture, or make) influenced the participants' speech during their description and whether those differences in speech affected memory for letters or locations in the grid. We used the lme4 package (Bates et al., 2015) in R (version 4.1.0) to predict each speech measure (word count, filled pauses, time spent speaking, and deictic reference) from the fixed factor of condition. We included random effects of participant and item[1] with the maximal random effects structure (intercepts and slopes by condition). If this maximal model did not converge, we dropped the random slopes. We determined significance by using the lmerTest package (Kuznetsova et al., 2017), which uses Satterthwaite's method for approximating degrees of freedom and determining the

---

[1] Note that the random effect of 'item' includes variability due to the particular pattern being described, the particular grid being remembered, and the ordinal position of the trial, as these were confounded in the design (i.e., trial one always showed a particular pattern paired with a particular grid).

p-value associated with the *t*-test of each effect, and the difflsmeans function to test pairwise differences between the three conditions.

Next, we considered whether each speech measure was associated with memory for the letters or locations in the grid. To do this, we entered the measure as a fixed factor in (1) a linear mixed model predicting number of letters recalled and (2) a linear mixed model predicting number of locations recalled. In both models, we included participant and item as random effects and determined significance with the lmerTest package (Kuznetsova et al., 2017).

Finally, we considered whether condition (Gesture, Hands Still, and Make) affected cognitive load. We conducted a separate series of linear mixed-effects models to predict memory for letters and memory for locations. In each series, we first fit a full model that included condition and all four speech variables as fixed factors as well as participant and item as random effects. We began with the maximal model possible, including random intercepts and slopes by condition for both participant and item, and dropped the random slopes if the maximal model did not converge. We then computed a set of 14 nested models, encompassing all combinations of one, two, or three speech variables in addition to condition, and each including the same fixed and random effects structure as the full model. We compared the models using the Akaike information criterion (AIC), with lower AIC values indicating better model fit. We then chose the best-fitting model and used the difflsmeans function in the lmerTest package to determine pairwise differences between conditions in that model.

The data for both experiments and the analysis code are available at https://osf.io/57cdw/?view_only=dd7a46754c274608a896549a2f38168e.

2.2. Results

2.2.1. Speech in the primary description task

Table 1 displays the means and standard errors of the four speech variables (word count, time in seconds, filled pauses, and deictic speech) in each of the three conditions. All four variables differed significantly by condition. Specifically, the descriptions in the Make condition contained significantly fewer words and contained significantly fewer filled pauses than the descriptions in the Gesture or Hands-Still condition. Descriptions in the Hands-Still condition resembled those in the Gesture condition in the number of words, but took significantly longer and contained more filled pauses. This suggests that speaking with hands still was somewhat more difficult than speaking with gesture on this task, as it took participants more time and effort (as indicated by filled pauses) to produce a comparable number of words when they kept their hands still. Deictic references (e.g., 'like this') occurred

Table 1. Means (SE) of the speech variables in the three conditions of Experiment 1

Speech variable	Gesture	Make	Hands Still
Word count	$37.30 (2.37)_a$	$30.83 (1.73)_b$	$37.29 (2.30)_a$
Seconds	$16.89 (1.04)_a$	$19.03 (1.05)_b$	$18.41 (1.05)_b$
Filled pauses	$1.39 (0.17)_a$	$1.00 (0.17)_b$	$1.60 (0.17)_c$
Deictic speech	$0.17 (0.03)_a$	$0.24 (0.04)_b$	$0.02 (0.009)_c$

Note. For each variable, means with a different subscript differ statistically from one another ($p < 0.05$).

Table 2. Models predicting memory of letters and locations from each speech variable in Experiment 1

Speech variable	Letters				Locations			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Word count	−0.003	.003	0.84	0.40	0.0003	.003	0.11	0.92
Seconds	−0.012	.007	1.76	0.08	−0.008	.006	1.33	0.19
Filled pauses	−0.013	.033	0.41	0.68	−0.033	.030	1.09	0.27
Deictic speech	0.151	.122	1.23	0.22	0.026	.012	2.24	0.03

most frequently in the Make condition and least frequently in the Hands-Still condition.

Performance on the memory task was unaffected by the speech variables for the most part. As shown in Table 2, word count, time in seconds, and filled pauses produced during the description were all unassociated with both verbal memory of the letters in the grid and spatial memory of the locations. Deictic references did not predict memory for the letters recalled but did predict memory for the locations, with participants remembering more locations on trials in which they produced a deictic reference in their speech ($M = 2.61$, $SE = 0.10$) than on trials in which they did not produce a deictic reference in speech ($M = 2.29$, $SE = 0.04$).

2.2.2. Verbal memory on the secondary task

The maximal model did not converge, so full and nested models were used that included random intercepts for participant and trial without random slopes. The best-fitting model was determined to be the model that included filled pauses and condition. Participants' ability to remember the identity of the letters they had seen in the grid was significantly affected by their manual behavior during the description task (see Fig. 2). Participants remembered significantly more letters when they gestured with their description ($M = 2.96$, $SE = 0.14$) than when they kept their hands still ($M = 2.73$, $SE = 0.14$), $\beta = .24$, $SE = .09$. $t = 2.53$, $p = 0.01$. When participants made the pattern they were describing ($M = 2.88$, $SE = 0.14$), their performance was intermediate to and not significantly different from their performance when they described with hands still, $\beta = .15$, $SE = .10$, $t = 1.54$, $p = 0.12$, or when they described with gesture, $\beta = .08$, $SE = .10$, $t = 0.82$, $p = 0.41$. It appears that participants experienced lower cognitive load when they gestured on the primary task, and this lightened load benefited their ability to maintain the verbal material from the grid. Although the best-fitting model included filled pauses, the effect of filled pauses was not significant, $\beta = −0.005$, $SE = .03$, $t = 0.14$, $p = 0.89$.

2.2.3. Spatial memory on the secondary task

Full and nested models were used that included random intercepts for participant and item without random slopes because the maximal model did not converge. The best-fitting model including condition was determined to be the model that included filled pauses and deictic reference. Deictic reference was significant, $\beta = .30$, $SE = .12$, $t = 2.46$, $p = 0.01$, such that participants recalled more locations from the grid when they produced a deictic reference during the description of a pattern than when they did not. The number of filled pauses did not predict memory for locations, $\beta = −.04$, $SE = .03$, $t = 1.20$, $p = 0.23$. Importantly, there were no significant effects of condition;
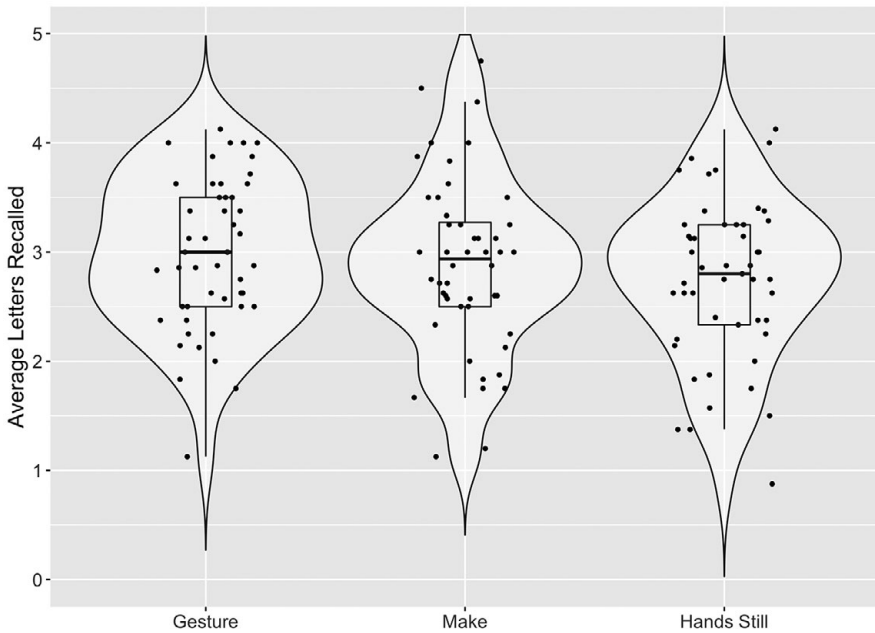
Figure 2. Average letters recalled on the secondary task as a function of manual activity during the description task in Experiment 1.

Note. The average number of letters recalled in the grid by the participants in Experiment 1 in each of the three conditions. Participants recalled more letters when they described with gesture than when they described with hands still.

participants' ability to remember the location of the letters they had seen in the grid was not significantly affected by their manual behavior during the description task. The number of locations recalled when participants described with gesture (M = 2.34, SE = .10) did not differ significantly from the number of locations recalled when they described with hands still (M = 2.38, SE = .10), $ß$ = .04, SE = .09, t = .42, p = 0.67, or when they described while making the pattern (M = 2.24, SE = .11), $ß$ = .10, SE = .09, t = 1.06, p = 0.29. There was also no difference in the number of locations recalled between describing with hands still and making the pattern, $ß$ = −.14, SE = .10, t = 1.40, p = 0.16.

2.3. Discussion

The results from the first experiment replicate the findings of previous studies showing that producing gestures uses fewer cognitive resources than keeping hands still (e.g., Goldin-Meadow et al., 2001). Further, we found that the benefit of gesture is not shared by action, as making the patterns as they were described did not produce a benefit to cognitive load compared to keeping one's hands still. Finally, we found that the cognitive benefit of gesture was manifest only on the verbal measure (how many letters they could recall) and not on the spatial measure (how many locations they could recall).

One limitation of Experiment 1 was that the patterns remained on the screen as participants described them. In this situation, it appears that there is little to be gained by making the pattern, as creating another visual representation of the pattern is not helpful when there is already one present on the screen. The fact that gestures reduced verbal load compared to keeping hands still in this situation suggests that the gestures were not helpful because they were making the information visible, as the information was visible on the screen in all conditions. Rather, gestures may have been helpful because they were helping the speaker focus on which aspects of the pattern were most important as they described it, and actions did not appear to produce this same benefit. Still, it remains unclear whether actions might benefit cognitive load in a situation where making a visible depiction of the pattern is advantageous, such as when describing the pattern from memory. In Experiment 2, we tested this possibility by repeating the method from Experiment 1, except that this time the pattern disappeared from the screen after 5 s and before participants were asked to describe it. We again measured participants' verbal and spatial load by considering the number of letters and locations that they were able to recall on the secondary task. We predicted that actions (i.e., making the pattern) may benefit cognitive load compared to keeping hands still.

3. Experiment 2

Participants again completed 24 trials in a within-subjects design. On each trial, they described a pattern while attempting to maintain spatial and verbal information on a secondary task. As in Experiment 1, they were instructed to gesture on eight trials, instructed to keep their hands still on eight trials, and instructed to make the pattern on eight trials. The only difference from Experiment 1 was that in Experiment 2, the pattern they were describing did not remain visible on the screen during their description. We again predicted that gesturing would be associated with lower cognitive load (higher performance on the secondary task) than keeping hands still. We also predicted that making the pattern would be associated with lower cognitive load (higher performance on the secondary task) than keeping hands still because making the pattern would provide an external visual representation of the pattern that would make it easier to describe. We again measured filled pauses, time spent speaking, words spoken, and the presence of deictic reference (e.g., 'like this') as measures of speaking difficulty.

3.1. Method

3.1.1. Participants

We again recruited participants from introductory psychology courses during a single academic term and allowed all students who wanted to participate the opportunity to do so. Thirty-eight (26 women; 12 men) produced usable data. All participants were between the ages of 18 and 20 years old, and the majority (84%) were first-year students. All participants included in the final sample were native English speakers, and the majority identified as white (60%), with Latinx (16%), Asian (8%), Black (8%), Multi-racial (5%), and Middle Eastern (3%) students also represented. Participants were compensated with extra credit in their Psychology course.

3.1.2. Procedure

All materials and procedures were identical to those used in Experiment 1, except that on each trial, the pattern to be described only remained on the screen for 5 s.[2] Once the pattern disappeared, the words 'MAKE', 'GESTURE', or 'DO NOT GESTURE' appeared on the screen in its place and participants were asked to describe the pattern they had seen while following the trial-specific instructions. After describing the pattern, they pressed a key and were then prompted to recall the information from the grid that preceded the pattern. As in Experiment 1, participants completed eight trials in each of the three conditions for a total of 24 trials, and the conditions appeared in a pseudo-random order so that participants were unaware when they were viewing the grid and the pattern which kinds of instructions would be given. Both verbal load (i.e., the number of letters recalled from the grid) and spatial load (i.e., the number of locations recalled) were measured on every trial.

3.1.3. Data exclusions and reliability

We excluded one trial in the Gesture condition because the participant did not describe the pattern. In the Hands-Still condition, 46 trials were excluded because the participant gestured at least once during the description, two were excluded because the participant inadvertently skipped the trials by pressing the space bar too many times, and one was excluded because the participant did not provide a description of the pattern. In the Make condition, 75 trials were excluded because the participant gestured at least one time in addition to making the pattern, and 30 were excluded because the participant did not describe the pattern as they were making it. An additional three trials were excluded because the participant inadvertently skipped them and two because the experimenter interrupted the trial to remind them of the instructions. These exclusions led to the loss of 17.5% of the data. The final dataset included an average of 7.97 trials per participant in the Gesture condition (Range: 7–8), 5.33 trials per participant in the Make condition (Range: 1–8),[3] and 6.71 trials per participant in the Hands-Still condition (Range: 4–8). The increase in trial exclusions compared with Experiment 1 is large because participants had a more difficult time inhibiting gesture on Hands-Still trials in Experiment 2 than in Experiment 1. This aligns with previous research that suggests that gestures are more prominent when speakers describe information from memory than when they describe information that is visually present (e.g., Wesp et al., 2001). In Experiment 2, the pattern was not present on the screen as participants described it, and participants had a more difficult time keeping their hands still.

We calculated reliability in the same way as in Experiment 1. Agreement between the two coders for whether a gesture occurred was again substantial (89% agreement; Cohen's k = .77). Exclusions were made on the basis of the first coder's judgments.

---

[2]Prior work with these patterns showed that 3 s was adequate for accurate encoding (Hostetter et al., 2007). However, we gave participants an additional 2 s to account for any increased difficulty due to the extraneous load they were under from trying to also remember the information from the grid.

[3]Two participants in Experiment 2 had only one trial remaining in the Make condition following exclusions, demonstrating a difficult time following instructions in this condition. The pattern of results described does not differ if these two participants are excluded from analyses.

3.2. Results

3.2.1. Speech in the primary description task

The average number of words produced, number of filled pauses produced, number of seconds taken on descriptions, and proportion of descriptions that contained deictic references (e.g., 'like this') in each condition are shown in Table 3. There were differences across conditions, such that descriptions in the Make condition contained fewer words and filled pauses than descriptions in the Gesture or Hands-Still condition. Descriptions in the Gesture condition contained fewer words and disfluencies than descriptions in the Hands-Still condition. As observed in Experiment 1, descriptions in the Hands-Still condition were significantly less likely to contain deictic references to the physical environment (e.g., 'it looks like this') than descriptions in the Make or Gesture condition.

Further, three of the speech variables significantly predicted the number of letters recalled on the secondary task (see Table 4). As the number of words, the number of filled pauses, and the duration spent on the description of the pattern increased, the number of letters successfully recalled on the secondary task decreased. However, using a deictic reference to the environment did not increase the number of letters or locations recalled on that trial.

3.2.2. Verbal memory on the secondary task

The full model did not converge with random slopes, so these were dropped from both the full and all nested models. The model with the best fit included seconds and condition. However, unlike in Experiment 1, gesturing with the description did not result in better memory for the letters on the secondary task ($M = 3.01$, $SE = 0.17$) compared to keeping hands still ($M = 2.89$, $SE = 0.17$), $ß = .12$, $SE = .11$, $t = 1.11$, $p = 0.27$ (see Fig. 3). However, in Experiment 2, making the pattern while describing it resulted in better memory for the letters ($M = 3.15$, $SD = 0.18$) than keeping hands

Table 3. Means (SE) of the speech variables in the three conditions of Experiment 2

Speech variables	Gesture	Make	Hands Still
Word count	31.52 (2.41)$_a$	24.54 (2.45)$_b$	33.57 (2.42)$_c$
Seconds	15.48 (1.23)$_a$	16.69 (1.26)$_b$	16.74 (1.24)$_b$
Filled pauses	1.49 (0.22)$_a$	1.05 (0.23)$_b$	1.84 (0.22)$_c$
Deictic speech	0.17 (0.04)$_a$	0.16 (0.05)$_a$	0.004 (0.004)$_b$

Note. For each variable, means with a different subscript differ statistically from one another ($p < 0.05$).

Table 4. Models predicting memory of letters and locations from each speech variable in Experiment 2

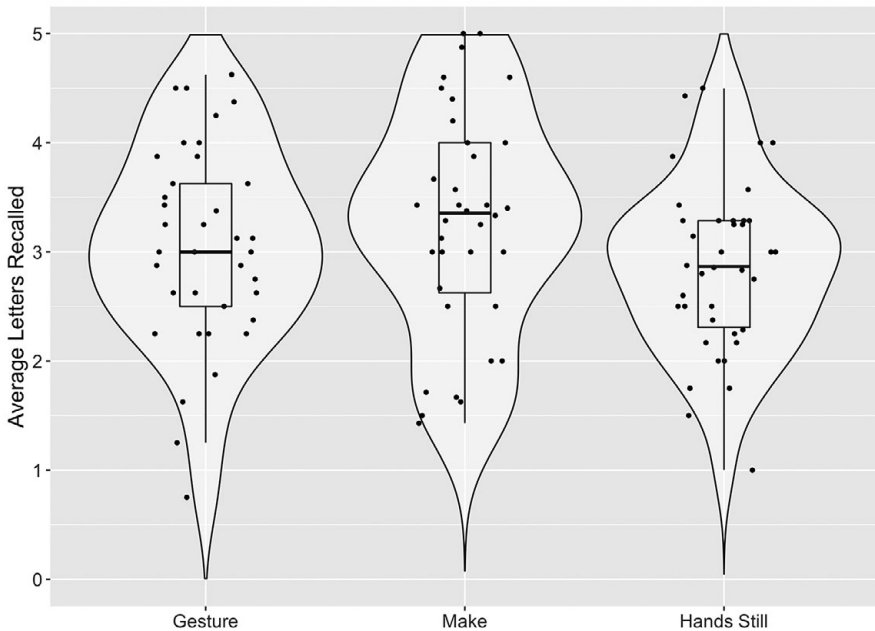| Speech variables	Letters				Locations			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Word count	−.015	.004	3.57	0.0003	−.006	.004	1.54	0.13
Seconds	−.028	.008	3.33	0.0009	−.009	.008	1.25	0.21
Filled pauses	−.10	.04	2.77	0.006	−.06	.03	1.87	0.06
Deictic speech	.21	.17	1.27	0.20	−.02	.15	0.11	0.91

Figure 3. Average letters recalled on the secondary task as a function of manual activity during the description task in Experiment 2.
Note. The average number of letters recalled in the grid by the participants in Experiment 2 in each of the three conditions. Participants recalled more letters when they made the pattern as they described it than when they described with hands still.

still, $\beta = .26$, $SE = .13$, $t = 2.06$, $p = 0.04$. There was no difference between making and gesturing, $\beta = .14$, $SE = .12$, $t = 1.13$, $p = 0.26$. It appears that when the pattern was not present on the screen, making the pattern during the description resulted in lower verbal load than keeping hands still. The best-fit model also included seconds, and this was a significant predictor of letters recalled, $\beta = -0.03$, $SE = .009$, $t = 3.27$, $p = 0.001$. As speakers spent longer describing a particular pattern, their memory for the letters in the grid decreased.

3.2.3. Spatial memory on the secondary task
Because the maximal model did not converge, random slopes were dropped from the full and nested models. The best-fit model included deictic speech in addition to condition, though deictic speech was not a significant predictor of locations recalled in Experiment 2, $\beta = -.02$, $SE = .16$, $t = 0.13$, $p = 0.90$. Most importantly, there was no significant effect of condition on locations recalled. Gesturing about the patterns did not result in reduced spatial load ($M = 2.39$, $SE = 0.13$) compared to making the patterns ($M = 2.53$, $SE = 0.14$), $\beta = .15$, $SE = .11$, $t = 1.35$, $p = 0.18$, or keeping hands still ($M = 2.44$, $SE = 0.14$), $\beta = .05$, $SE = .10$, $t = 0.46$, $p = 0.65$. There was also no difference between making the patterns and keeping hands still, $\beta = .10$, $SE = .12$, $t = 0.87$, $p = 0.38$. As in Experiment 1, manual behavior during the description task did not affect participants' memory for the locations on the secondary task.

3.3. Discussion

Experiment 2 did not replicate the finding from Experiment 1 that gestures reduced verbal load during a description task compared to describing with hands still. Instead, in Experiment 2, participants had the lowest verbal load (i.e., were able to remember the most verbal information on the secondary task) when they made the pattern as they described it. Participants' verbal behavior was also different when they made the pattern, as they used fewer words and filled pauses than when they gestured or kept their hands still, which suggests that describing the pattern was easier when speakers were making a visual representation of the pattern that they could reference. Further, these verbal differences were predictive of success with recalling the verbal information from the grid in Experiment 2. It appears that when speakers can see what they are describing, the burden of speaking is reduced, and this lightened load frees up resources that can be used to remember the secondary information from the grid.

4. General discussion

In two experiments, we compared the cognitive load involved in describing with gesture to the cognitive load involved in describing with action or describing with hands still. We found that, compared with keeping hands still, gestures reduced verbal load when the pattern remained visible on the screen (Experiment 1). When the pattern was not visible on the screen, making the pattern reduced verbal load compared with keeping hands still (Experient 2). However, there was no difference in the cognitive load associated with gesture and action in either experiment. This suggests that moving the hands in meaningful ways while speaking reduces cognitive load and that the similarities between gesture and action are perhaps more important than their differences. The similarity in gesture and action makes sense under some views of gesture which propose that action and gesture originate from the same underlying processes (e.g., Chu & Kita, 2016; Hostetter & Alibali, 2008, 2019).

Despite the similarity of gesture and action in the present experiments, we did find that they seemed to benefit cognitive load in different situations and likely for different reasons. Gesture benefited verbal load over keeping hands still when the pattern remained visible during description (in Experiment 1) but not when the pattern was no longer visible (in Experiment 2). This suggests that gesture's effect on cognitive load may primarily come from its attention-directing function – by helping the speaker keep track of where they are in a complex, visually present pattern (see also Pouw et al., 2016). For example, by pointing to a particular dot or shape on the computer screen, speakers can index their thinking to the physical environment rather than having to internally maintain the entirety of the pattern in their working memory. This can then help a speaker who has finished describing one element of the pattern more easily discern where they were and what elements still need to be explained. Note that under this view, gesturing does not change verbal output in very noticeable ways, which aligns with our finding in Experiment 1 that the speech on gesture and hands-still trials was comparable in terms of filled pauses and the number of words spoken, but that the amount of time spent speaking was shorter on gesture trials, as we would expect if gestures are helping speakers make decisions about what to say more quickly. In contrast, we did not find evidence for the possible benefit of using gesture to recreate a pattern that is not visually present and keep it active in their working memory (e.g., Wesp et al., 2001).

Although speakers may use gesture in this way, this function of gesture does not seem to reduce cognitive load on a secondary task, as in Experiment 2 we did not find a cognitive benefit of gesture compared to keeping hands still when the pattern was not visually present.

This finding differs from that reported by Ping and Goldin-Meadow (2010), who found a cognitive benefit of gesture regardless of whether the information being described was visually present. Several differences between the present study and Ping and Goldin-Meadow should be noted. First, the stimuli used by Ping and Goldin-Meadow were objects in a Piagetian conservation task (e.g., cups of water). These stimuli are arguably less complex than the spatially rich stimuli used here, which included difficult-to-name shapes (e.g., rhombus and parallelogram) and a need to describe precise spatial relationships among multiple components (e.g., above; to the left of; at a 45-degree angle). Second, the explanation task examined by Ping and Goldin-Meadow was highly conceptual (e.g., explain why these two amounts of water are the same or different), whereas the explanation task used here was highly descriptive (e.g., describe the location of the dots in the pattern you saw). Finally, the participants in Ping and Goldin-Meadow were children, whereas the participants in the present sample were college-aged adults. Any of these differences could have resulted in the discrepancy in findings across the two studies.

Interestingly, we saw the opposite pattern of benefit for action – making the pattern benefited cognitive load compared to keeping hands still only when the pattern was not visible (Experiment 2). This suggests that creating a pattern through action that leaves a visual trace is helpful to cognitive load, perhaps by reducing the need to keep the entire pattern in one's working memory, by helping speakers see the relationship among different components of the pattern, and by keeping track of how much has been described and what still needs to be described. As a speaker begins describing the first part of a pattern that they are also making, they can visually see the components they are describing. As they progress in their description, they can more easily see how the next part relates to the first part they already described and made, which is visually present on the table. This could result in an easier time with conceptual planning (e.g., the speaker can see that the four dots they placed on the table make a square or that the next dot goes out and to the left) or with lexical access (e.g., by seeing a square in front of them, the corresponding word is more highly activated). It could also help the speaker more readily keep track of the fact that four dots have been used and there are three left, for example. Regardless of which function is paramount, making the pattern seems to reduce the verbal load involved in describing compared to keeping hands still.

It is unclear from the present data whether gesture or action had any long-term effects on memory for the information described in the primary task. Some previous work has shown that gesture leads to better memory for the gestured information than keeping hands still (Cook et al., 2010) or than producing action (So et al., 2014), but we did not measure memory for the spatial patterns that were described in the present study. However, what is clear from the present work is that there is no extra cognitive *cost* associated with producing a gesture compared to producing an action, as in neither experiment did gesture result in significantly worse secondary memory performance than action. This suggests that any long-term benefits of one over the other may not come with short-term costs, as would be expected if gesture were

similar to other 'desirable difficulties' (Bjork & Bjork, 2011) that are difficult in the moment but come with long-term benefits.

In sum, it appears that both gesture and action can benefit a speaker's cognitive load. Gesture benefits load when it accompanies a visually present pattern, perhaps by directing participants' attention to the various components of the pattern as they are described, whereas action benefits load when it accompanies a pattern that is not visually present, perhaps by creating a visual trace that can then be referred to and thought about more easily during the description. Of course, in a more naturalistic setting, gesture and action are probably not as separable as they were here. We discouraged participants from also gesturing in the Action condition and discarded trials in which they did so in order to isolate the effects of gesture and action; however, in the real world, a speaker who made the pattern might then gesture to it, which could further benefit their cognitive load.

It is also worth noting that the benefits of gesture and action in the present experiments were manifest on a secondary verbal task but not spatial task. Although most investigations of the effect of gesture on cognitive load have examined verbal load (e.g., Marstaller & Burianova, 2013), the only previous study to include measures of both verbal and spatial load found evidence that gestures benefited both types of loads (Wagner et al., 2004). One notable difference in our study compared with Wagner et al. could have produced this discrepancy. In the present study, participants attempted to remember both verbal and spatial information on every trial, requiring participants to prioritize which type of information to focus on. We suspect that participants may have prioritized the verbal information because it felt somewhat easier to manage on top of the highly spatial primary task. If participants simply did not try very hard to rehearse or retain the spatial information from the grid, then there may not have been any effort for gesture or action to affect. More studies are needed that specifically examine how gestures might affect memory for spatial information in a secondary task, as to date, almost all studies in this program of research have examined verbal information (with Wagner et al., 2004 as the one notable exception). Until then, caution is warranted in interpreting the present null effect as evidence that gesture does not benefit spatial load.

These questions could be further explored by considering the role of individual differences in these effects. There are known to be individual differences in how much speakers spontaneously gesture (e.g., Chu et al., 2014), and the effects of gesture for cognition may be different for those with lower versus higher working memory capacity (e.g., Pouw et al., 2016). Indeed, Marstaller and Burianova (2013) found that the benefit of gesture for verbal working memory was strongest for participants with low verbal working memory. Perhaps the benefits for spatial working memory are also strongest for participants with low spatial working memory. We cannot address these questions in our sample because we did not measure participants' spatial (or verbal) working memory in a way that was independent of the task. In future work, it might be worth specifically recruiting samples that are thought to have low (or high) spatial working memory or assessing the working memory of participants with a separate measure before the task begins.

Finally, the sample size and number of trials per participant were limited in the present experiment. In designing this study, we were mindful of how cognitively demanding our task was likely to be for participants and intentionally kept the duration of the study to be under 30 min. Because we also wanted to get data from

participants in all three conditions (Gesture, Action, and Hands Still), we only asked each participant to do eight trials in each condition. Combined with the fact that some of these trials were then excluded due to mistakes in following the instructions, our final analyses may have been underpowered to detect some potential effects. For example, with a larger dataset, perhaps a difference between the Gesture and Hands-Still conditions would have emerged in Experiment 2. Future studies would benefit from either including a larger total number of participants or from including only two of the three conditions in order to get more data per condition from each participant.

4.1. Conclusion

The present study joins an established body of literature (e.g., Goldin-Meadow et al., 2001; Marstaller & Burianova, 2013; Ping & Goldin-Meadow, 2010; Wagner et al., 2004) suggesting that gesture production can relieve the cognitive load involved in speaking. Importantly, while previous work has primarily examined the effect of gesture during explanations of math problems, the present experiments demonstrate that the effect also occurs when speakers are describing highly spatial information. Further, we have shown that action can also reduce load, particularly when the information being described is not visually present, and that the benefits of gesture and action are most readily seen on verbal (rather than spatial) load. Together, these findings both extend what is known about how gestures affect cognitive load of the speaker and provide more nuance to our understanding of the effect. Speakers who move their hands in meaningful ways reap benefits that reduce the overall verbal load of the task.

Data availability statement. The transcribed, anonymized datasets for both experiments described in this manuscript are publicly available at https://osf.io/57cdw/?view_only=dd7a46754c274608a896549a2f38168e. The analysis code for all analyses described in this manuscript is also available.

Competing interest. The authors have no known competing financial interests or personal affiliations that could have appeared to influence the work reported in this paper.

References

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, 130(4), 553–573. https://doi.org/10.1037/0033-299.130.4.553

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

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 42, 257–270. https://doi.org/10.1037/a0021790

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. https://doi.org/10.1037/a0033861

Church, R. B., Kelly, S., & Holcombe, D. (2014). Temporal synchrony between speech, action, and gesture during language production. *Language, Cognition, and Neuroscience*, 29(3), 345–354. https://doi.org/10.1080/01690965.2013.857783

Cohen J. D., MacWhinney, B., Flatt M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257–271.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gestures make memories that last. *Journal of Memory and Language*, 63(4), 465–475. https://doi.org/10.1016/j.jml.2010.07.002

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language, Cognition, and Neuroscience*, 27(4), 594–610. https://doi.org/10.1080/01690965.2011.567074

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. https://doi.org/10.1146/annurev-psych-113011-143802

Goldin-Meadow, S., & Beilock, S. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5, 664–674. https://doi.org/10.1177/1745691610388764

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

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

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. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, 26, 721–752. https://doi.org/10.3758/s13423-018-1548-0

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language, Cognition, and Neuroscience*, 22(3), 313–336. https://doi.org/10.1080/0169096060600632812

Hostetter, A. B., Pouw, W., & Wakefield, E. M. (2020). Learning from gesture and action: An investigation of memory for where objects went and how they got there. *Cognitive Science*, 44, e12889. https://doi.org/10.1111/cogs.12889

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience*, 139(1), 39–50. https://doi.org/10.1016/j.neuroscience.2005.07.002

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture and action during language comprehension. *Psychonomic Bulletin & Review*, 22, 517–523. https://doi.org/10.3758/s13423-014-0681-7

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

Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes*, 24, 761–775. https://doi.org/10.1080/01690960802327971

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7, 54–60. https://doi.org/10.1111/1467-8721.ep13175642

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82 (13), 1–26. https://doi.org/10.18637/jss.v082.i13

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

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

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin and Review*, 24, 652–665. https://doi.org/10.3758/s13423-016-1145-z

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition*, 146, 339–348. https://doi.org/10.1016/j.cognition.2015.10.014

Ping, R. M., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, 34, 602–619. https://doi.org/10.1111/j.1551-6709.2010.0112.x

Pouw, W. T., Mavilidi, M. F., van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing: International Quarterly of Cognitive Science*, 17(3), 269–277. http://doi.org/10.0.7/s10339-016-0757-6

So, W. C., Ching, T. H.-W., Lim, P. W., Cheng, X., & Ip, K. Y. (2014). Producing gestures facilitates route learning. *PLoS One*, 9(11), e112543. https://doi.org/10.1371/journal.pone.0112543

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

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

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, 114 (4), 591–600. https://doi.org/10.2307/1423612