Acta Psychologica 197 (2019) 131–142

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

The role of gesture as simulated action in reinterpretation of mental imagery

Kevin L. Kamermans[a,*,1], Wim Pouw[a,b,1], Luisa Fassi[a], Asimina Aslanidou[a], Fred Paas[a,c], Autumn B. Hostetter[d]

[a] Erasmus University Rotterdam, Department of Psychology, Education and Child Studies, the Netherlands
[b] University of Connecticut, Department of Psychological Sciences, USA
[c] University of Wollongong, School of Education/Early Start, Australia
[d] Kalamazoo College, Psychology Department, USA

ARTICLE INFO

Keywords:
Gesture
Haptic perception
Manual interference
Mental imagery
Reinterpretation
Simulated action

ABSTRACT

In two experiments, we examined the role of gesture in reinterpreting a mental image. In Experiment 1, we found that participants gestured more about a figure they had learned through manual exploration than about a figure they had learned through vision. This supports claims that gestures emerge from the activation of perception-relevant actions during mental imagery. In Experiment 2, we investigated whether such gestures have a causal role in affecting the quality of mental imagery. Participants were randomly assigned to gesture, not gesture, or engage in a manual interference task as they attempted to reinterpret a figure they had learned through manual exploration. We found that manual interference significantly impaired participants' success on the task. Taken together, these results suggest that gestures reflect mental imaginings of interactions with a mental image and that these imaginings are critically important for mental manipulation and reinterpretation of that image. However, our results suggest that enacting the imagined movements in gesture is not critically important on this particular task.

When speakers describe mental transformations, spatial scenes, or movement, they often move their hands in a way that depicts some of the spatial and kinematic characteristics of what they are describing. Cognitive scientists have become increasingly interested in how such gestures relate to speakers' mental imagery (e.g., Chu & Kita, 2011; Hegarty, Mayer, Kriz, & Keehner, 2005; Hostetter & Alibali, 2008; Pouw, Mavilidi, Van Gog, & Paas, 2016). Two questions have been at the heart of this research. First, to what extent do gestures *reflect* a speaker's mental imagery? That is, are gestures created for the communicative act of describing or are they tied more fundamentally to the speaker's mental imagery? Second, to what extent do gestures *affect* a speaker's mental imagery performance? That is, are gestures beneficially involved in the creation, maintenance, or transformation of a mental image? We examine these questions in two studies by examining the frequency (Experiment 1) and the function (Experiment 2) of gestures in a task that requires participants to reinterpret mental images.

As defined by Kosslyn, Thompson, and Ganis (2006), "A mental image occurs when a representation of the type created during the initial stages of perception is present but the stimulus is not actually being perceived" (p. 122). Although mental imagery can correspond to any of the senses (e.g., olfactory, auditory), *visual* imagery has received the most attention from cognitive scientists. Some have questioned whether the experience of visual imagery can be explained by top-down propositional encodings without appealing to structural resemblance with perception (e.g., Pylyshyn, 2002). However, it is now clear that the same neural systems involved in early vision are also involved in visual imagery (e.g., Slotnick, Thompson, & Kosslyn, 2005; for an overview see Pearson & Kosslyn, 2015) and that sensory and motor processes are recruited during mental imagery. For example, during visual imagery, participants move their eyes in spatial synchrony with visually imagined scenes (e.g., Brandt & Stark, 1997; Laeng & Teodorescu, 2002; Spivey & Geng, 2001), even when they keep their eyes closed (e.g., Spivey, Tyler, Richardson, & Young, 2000). These eye movements do not give access to "visual information" in a classic sense (i.e., they provide no "input"), yet it seems that this enactive bodily movement is *functionally* involved in visual imagery processes (e.g., Johansson & Johansson, 2014; Laeng, Bloem, D'Ascenzo, & Tommasi, 2014; Pearson & Logie, 2015; Thomas, 1999).

In addition to moving their eyes as though viewing an imagined scene, people also sometimes move their hands as though interacting with the objects they imagine. For example, during a mental transformation task in which an object must be imagined as it would appear

---

[*] Corresponding author at: Department of Social Sciences, Erasmus University Rotterdam, 3000 DR Rotterdam, the Netherlands.
E-mail address: kamermans@essb.eur.nl (K.L. Kamermans).

[1] Kevin L. Kamermans and Wim Pouw share first authorship.

https://doi.org/10.1016/j.actpsy.2019.05.004
Received 21 September 2018; Received in revised form 19 March 2019; Accepted 8 May 2019

after being rotated around its axis (e.g., Shepard & Metzler, 1971), people often gesture as though they are physically rotating the object (e.g., Chu & Kita, 2008, 2011), and they gesture more than they do when describing the same object in its static end state (Hostetter, Alibali, & Bartholomew, 2011). Indeed, one view of gestures is that they emerge from simulations of action and perception states that are activated in the interest of mental imagery (Hostetter & Alibali, 2008, 2018). As speakers think about or describe events, their motor systems become activated as though actually experiencing those events, and some of the motor activity comes to be expressed as gesture.

This view, termed the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008), finds empirical support from studies that have manipulated the amount of motor involvement in a task and observed corresponding changes in gesture rate. For example, several studies have found that speakers gesture more as they describe objects that are highly manipulable than as they describe objects that are not manipulable (Hostetter, 2014; Masson-Carro, Goudbeek, & Krahmer, 2016, 2017; Pine, Gurney, & Fletcher, 2010). Further, Hostetter and Alibali (2010) found that speakers gestured more as they described a pattern they had actually made than as they described the same pattern that they had only viewed. Chu and Kita (2016) found that participants gestured more as they thought about how to rotate a mug that was perceived as more graspable than as they imagined how to rotate a mug that was not perceived as graspable because it had spikes along its handle. Thus, it appears that gestures are more prevalent when people are actively imagining how they would physically interact with the objects they speak and think about. Under the view of the GSA framework, gestures reflect motor components of mental imagery that are activated with enough strength such that they spill over into actual motor execution.

Actualizing the motor components of imagery in gesture may have beneficial effects for the gesturer's own cognition (Ehrlich, Levine, & Goldin-Meadow, 2006; Novack & Goldin-Meadow, 2017; Pouw & Hostetter, 2016; Smithson & Nicoladis, 2014). Indeed, converging evidence suggests that expressing information in gesture might be an important causal agent in the formation of new problem representations (e.g., Boncoddo, Dixon, & Kelly, 2010; Stieff, Lira, & Scopelitis, 2016). For example, Boncoddo et al. (2010) found that children who gestured about how gears turned in alternating directions were more apt to discover a higher-order rule that the turning direction of interlocking gears alternates than children who did not gesture. It appears that gestures may lead to the discovery of new information by externalizing and actualising mental imagery (Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). More specifically, gestures seem particularly apt for facilitating thinking about mental transformations to an image. For example, in mental rotation problems, solvers must learn how changes in one component of the image result in changes to the other components. In these tasks, participants are given a starting state and instructions about how to transform the image (which they must mentally imagine), and then asked to choose which given picture would correspond to the end state following the transformation. Gestures occur with high frequency during these types of tasks (e.g., Hegarty et al., 2005), and they increase in frequency as the difficulty of the rotation increases (Chu & Kita, 2011). Chu and Kita (2011) found that participants who were encouraged to gesture (or spontaneously gestured) on a mental rotation task outperformed those who did not gesture—not only on the mental rotation task, but also on a subsequent paper-folding task that required similar spatial transformation processes. This effect occurred even though no participants gestured on the paper-folding task. Chu and Kita proposed that the act of gesturing on the mental rotation problems helped participants learn the process of spatial transformation, which they could then also apply to the paper-folding task. Thus, it appears that externalizing the rotation process in gesture provides an intermittent scaffolding that eases the cognitive burden involved in learning to visualize a transformation internally.

What is it about producing a gesture that helps the producer better understand mental rotation? One possibility is that the production of a gesture creates a visible trace that can then be observed perceptually. Speakers gesture more when describing a scene that is not visually present than when describing the same scene visually before them, and it has been suggested that gestures are a way of keeping an image active in short-term memory (Wesp, Hesse, Keutmann, & Wheaton, 2001) or perceptually active (Pouw et al., 2014). However, visual perception cannot be the entire explanation, as it has been found that producing a gesture has greater effects for mental rotation than seeing someone else gesture, which also creates a perceivable visual trace (Goldin-Meadow et al., 2012). Thus, there seems to be something specific about the *kinesthetic* perceptual experience of producing a gesture that is helpful in thinking about mental rotation.

Previous research investigating the functional role of gesture in mental imagery has used tasks that primarily rely on spatial imagery. Spatial imagery, which concerns relative locations in space, can be contrasted with visual imagery, which concerns visual characteristics of an object such as shape or color (Kosslyn et al., 2006). Although a particular mental image can have both visual and spatial characteristics as defined above, there is evidence to suggest that visual and spatial components can be differentially activated (Kozhevnikov, Kosslyn, & Shephard, 2005). Further, the connection between movement and spatial imagery has been well supported empirically. For example, inhibiting eye movements impairs performance on a spatial working memory task (Postle, Idzikowski, Sala, Logie, & Baddeley, 2006), and eye movements are particularly likely to be involved when the spatial components of a visual image need to be realized (Sima, Schultheis, & Barkowsky, 2013). In tasks such as mental rotation, it is arguably the spatial components of the image that are most relevant to the task (Linn & Petersen, 1985). That is, participants must track where the various components of a shape are in space relative to some frame of reference (e.g., the vantage point of the body) as the motion of rotation occurs. Thus, the finding that gestures assist with mental rotation fits with our general understanding of how movement is integral to the formation and maintenance of spatial imagery.

It is possible that gestures may also support visual components of imagery. Some have suggested that sensorimotor systems coordinate during imagination (Foglia & O'Regan, 2016; Pearson & Kosslyn, 2015; Thomas, 1999). Rather than operating strictly on "separate formats for motor, auditory, kinesthetic, and tactile information" (Pearson & Kosslyn, 2015, p. 10091), sensorimotor systems may work in concert to bring forth imagery (Foglia & O'Regan, 2016; Thomas, 1999). Under this view, simulating the actions that were involved in perception (e.g., eye movements, manual interaction) does not just reactivate the experience of perception in that particular modality, but can also strengthen imagery cross-modally. Under such a view, imagining how the contours of an object would look involves knowing how that object would feel, and this sensorimotor contingency between seeing and feeling is partially expressed in gesture (or for example eye movements). Expressing these movements in gesture may thus inform the entire multimodal experience of imagery, as externally simulating how the object's contours feel could also inform how they would *look*, i.e., inform about the visual form of the object (next to its relative orientation to some frame of reference).

There is some evidence that movement could support visual imagery in this way. When people tactically perceive objects, movements that trace the contours of the object are critically important for identifying the object's exact shape (Lederman & Klatzky, 1993). It is possible that simulation of these same movements is important during visual imagery of the object's shape. Furthermore, it has been demonstrated that visual training to recognize shape categories of objects is transferable and does not only lead to improvement in visual performance, but also in performance on haptic recognition (Wallraven, Bülthoff, Waterkamp, van Dam, & Gaißert, 2014). This effect works vice versa, meaning that training through the haptic modality also transfers to increased performance in the visual modality. These results suggest that object shape can be brought forth in imagery through either (imagined) touch or through (imagined) vision, and that both streams of information are directly co-informative and coherent. This is further corroborated by neural evidence demonstrating that posterior inferior temporal regions of the visual cortex are activated when participants perceive an object through touch as well as through vision (Pietrini et al., 2004). Finally, perception of an object's shape has been shown to activate the dorsal visual stream, which is fundamentally involved in spatial and motor processing (Oliver & Thompson-Schill, 2003). Taken together, such evidence lends support to the hypothesis that simulating manual exploration of an object could strengthen the visual imagining of the object's shape.

Some evidence suggests that *gestures* might support visual imagery in this way. Specifically, individuals who have low visual working memory capacity gesture more than those with high capacity, and visual working memory capacity seems to be a better predictor of how much individuals gesture than spatial working memory capacity (Chu, Meyer, Foulkes, & Kita, 2014a, 2014b). Further, visual working memory (but not spatial working memory) capacity predicts how much individuals are helped by gesture on the Tower of Hanoi task (Eielts et al., 2018; see also Pouw et al., 2016). Thus, individuals with low visual working memory are particularly likely to both produce and be helped by gestures, as would be expected if gestures facilitate the formation or maintenance of visual images.

In the present study, we consider whether gesture might play a functional role in a task where the *visual* characteristics of the image are of primary importance. Toward this aim, we employed a classic paradigm in visual imagery research in which participants attempt to reinterpret a memorized figure (Finke, Pinker, & Farah, 1989; Mast & Kosslyn, 2002; Peterson, Kihlstrom, Rose, & Glisky, 1992). These figures are bistable, meaning that they have alternative interpretations depending on the orientation in which the figure is presented (see Fig. 1). To be successful at reinterpreting such figures, participants must successfully mentally reinspect it to subsequently “see” the alternative interpretation. Kamermans, Pouw, Mast, and Paas (2017) showed that, although difficult, about 30% of participants who learned the figure in one orientation could mentally rotate it and perceive the alternative interpretation using mental imagery. We believe this task is interesting for examining the role of gesture in imagery because success on the task largely depends on visual imagery. Although spatial imagery is involved to mentally rotate the figure, the ability to “see” the alternative interpretation in the rotated shape depends on participants' ability to holistically perceive the end state of the rotation, rather than compare its spatial components to those of a visually presented stimulus as in traditional mental rotation tasks.

In Experiment 1, participants were exposed to two bistable figures in one of their orientations and asked to memorize them, one by using vision-only and one by using touch-only to feel the contours of the shape. Participants were then introduced to the idea of bistability and asked to find the alternative interpretation of each figure by mentally rotating it. We predicted that, if gestures emerge from simulations of the actions that were involved during perception (Hostetter & Alibali, 2008), then participants should gesture more when explaining the alternative interpretation of the figure they had learned via touch-only than when explaining the alternative interpretation of the figure they had learned via vision-only. Further, we predicted that if gestures serve a beneficial role in the visual experience of imagery, then gesturing should be associated with more success finding the alternative interpretation of the figure than not gesturing, particularly in the touch-only condition where manual movements were paramount to the original perception of the figure. This is because simulating the movements involved during the actual perception of the figure might strengthen the visual image of the figure by bringing forth bodily co-regularities that are informative for the figure's shape, thereby making it easier to “see” the alternative interpretation (Pouw & Hostetter, 2016; Thomas, 1999).

1. Method Experiment 1

1.1. Participants & design

Sixty-eight students participated in Experiment 1 (89% female, *M*age = 20.30 years, *SD*age = 3.36, range 17–37 years). We recruited Dutch (83.8%) and Non-Dutch students from the Erasmus University Rotterdam who were all enrolled in Psychology Bachelor or Master programs (English spoken program) and participated for study credits. All participants were instructed and tested in English. The experiment used a within-subjects design with perceptual modality (visual-alone vs. touch-only) as the independent variable and reinterpretation performance (correct vs. incorrect) and gesture rates as the main dependent variables. The experiment consisted of two phases: A perception phase in which participants learned two test figures (one through vision and one through touch) and an imagery phase in which they had to mentally reinterpret the same figures they learned during the perception phase.

1.2. Materials

1.2.1. Test figures and equipment

Two bistable figures were cut from high-density foam sheets (thickness = 0.5 cm, length = ca. 16 cm, width = ca. 21 cm). As shown in Fig. 1, each figure represented the body of an animal in the 0 degrees orientation, and the head of a different animal in the 180 degrees orientation. The figures were originally designed by Leo Burnett Worldwide (2015) for a marketing campaign and were validated for their use in a mental reinterpretation task by Kamermans et al. (2017).

In the touch-only condition, we attached the figure with Velcro-tape to a small wooden base (height = 10 cm, width at the top = 5 cm) and placed base and figure under a closed cardboard box with two slot openings for participants' hands. In this way, visibility of the figure was blocked, but participants could feel the contours of the figure by putting their hands in the box. Because the figure was attached to a base, it could not be picked up, moved, or manually rotated during the haptic exploration of the figure. However, the width of the base was smaller than the figure's width at any point, so the participants could feel the edges of the shape distinctly without touching the base.

1.2.2. Questionnaire

Age, sex, and native language were reported by the participants.

Fig. 1. Line drawings of the test figures that were used. On the left seal/deer representation, and on the right penguin/giraffe.

Additionally, participants answered the following questions: "What do you think was the purpose of the current study? (If you have no idea, no answer is necessary)", and "What do you think the researchers are expecting to discover with the current study? (If you have no idea, no answer is necessary)". Five participants gave a correct description of the purpose of the experiment. However, these participants were not excluded from analysis as knowing the purpose of this study would not have improved their aptitude to successfully find the correct novel interpretation (akin to how knowing the purpose of an IQ test is not likely to increase your intellectual aptitude).

1.3. Procedure

All participants were tested individually and were told that they would participate in a study about imagination. Participants were video-recorded for the total duration of the experiment. The experiment consisted of a perception phase and an imagery phase, and participants completed one trial in each of two conditions (vision-only; touch-only) in each phase.

1.3.1. Perception phase

In the perception phase, participants inspected the two figures one at a time for 30 s in order to memorize them. One figure was inspected by visually viewing the figure (vision-only condition), and the other figure was inspected by touching the contours of the figure (touch-only condition). In the touch-only condition, participants put their hands through the slots of the cardboard box and manually inspected the figure for 30 s. In the vision-only condition, participants were shown the figure for 30 s and were not allowed to touch it. In both conditions, participants were instructed to form an accurate memory of each figure as they would be tested for their memory later on. The order of conditions (vision-only vs. touch-only), and figure assignment (deer/seal vs. giraffe/penguin) per condition were counterbalanced, as was the orientation (i.e., body vs. head orientation) that the bistable figure was presented in.

After 30 s of perceiving the figure either by vision-only or by touch-only, participants reported what they had seen or felt (depending on condition). If they reported two or more distinct interpretations of the figure during perception (i.e., premature bistability detection), they were not asked to reinterpret that figure in imagery (see Data exclusions below).

1.3.2. Imagery phase

Following the perception of both figures, participants were instructed to close their eyes and bring back a particular figure in mental imagery. The particular figure was referred to by the interpretation the participant provided in the perception phase. For example, when a participant reported perceiving a "cow" in the memorization phase, the experimenter would ask the participant in the testing phase to bring back their memory of the "cow". Once participants stated that they had retrieved a memory of the figure, they were informed that the figure had another interpretation that could be detected when rotating the figure 180 degrees, i.e., upside down. They were then asked to give their best guess for what the figure would be in its alternate orientation. The same process was then repeated for the second figure.

1.4. Performance scoring

Performance was measured as a dichotomous variable (no [correct] second interpretation vs. correct second interpretation). We followed the procedure used by Kamermans et al. (2017) to determine whether an interpretation was correct or not. Specifically, we compared each response given in the imagery phase to the set of responses given in the perception phase for that figure in that orientation and modality. If at least one person had given the interpretation during the perception phase with that modality, we scored the interpretation as correct. For example, if a participant reported a "cow" during the imagery phase as an alternate interpretation for a figure learned via touch-only, this was coded as correct if another participant who was presented with that figure in that orientation during the touch-only perception phase reported perceiving it as a cow. In this way, we considered any interpretation that was offered by participants during perception as a valid interpretation of the stimuli. Similar to other studies using this paradigm (e.g., Chambers & Reisberg, 1985; Mast & Kosslyn, 2002; Peterson et al., 1992), a limited number of post-hoc experimenter decisions were made to count answers as correct that were not mentioned in the corresponding perception phase, but that shared the same objective features as the target figure (e.g., reindeer head or impala were both counted as correct answers for the deer figure). Frequency of specific (in)correct interpretations can be retrieved from the Open Science Framework https://osf.io/725te/.

Note that this is a conservative measure of accuracy, as it primarily considers answers that were given by another participant in the same modality. Thus, we did additional performance scoring based on whether an answer was given during the perception phase in *any* modality. For example, if a participant in the imagery phase of a haptic trial said "airplane," this was coded as correct if it was offered by a participant during the perception phase for that figure regardless of modality. The current study presents the analyses based on the former more conservative measure of success with the reinterpretation task—the one based on responses given during perception only with the congruent modality. However, analyses using the latter more liberal coding scheme considering responses regardless of perceptual modality yield the same conclusions. Complete details of these additional analyses can be found at https://osf.io/725te/.

We did not count as correct any reinterpretations that belonged to the non-rotated interpretation (for example "upside down Eiffel tower"). Additionally, as in our previous research (Kamermans et al., 2017), we did not count letter symbols as correct because they are considered very simple and highly memorized symbolic representations that do not compare to iconic mental representations (cf. Finke et al., 1989).

1.5. Speech and gesture coding

We measured the amount of time and the number of words participants used as they described their interpretation on each trial. We also coded the amount of time they spent gesturing during each description. To do this, we timed from the moment the hands began moving until they came back to rest, and if multiple distinct gestures were performed in a trial, we summed the times for each gesture. A second rater coded 20% of the time measurements to calculate the interrater reliability using the intraclass correlation coefficient (ICC). Agreement for amount of time on each trial was good, ICC = 0.788, 95% CI [0.535, 0.904], $p < .001$. Agreement for amount of time participants spent gesturing was excellent, ICC = 0.838, 95% CI [0.628, 0.928], $p < .001$.

Note that participants who gestured often produced many interconnected movements that were difficult to segment into distinct gestures. For this reason, we did not count the number of gestures produced. However, for each participant who gestured, we noted the motion type of the gesture they produced. Specifically, whether they produced a *rotation* movement (defined as a grasp-handshape accompanied by a rotation of the wrist, as though showing how the figure would rotate), a *tracing* movement (defined as a sequential movement that appeared to be tracing the contours of the shape), or *other* (defined as any other movement). Fig. 2 shows an example for each type of movement produced. Participants who gestured could produce any or all of these movements. The same rater also coded 20% of the trials to calculate the interrater reliability using Cohen's Kappa. There was moderate agreement for *rotation* movements, $\kappa = 0.598$ ($p = .002$), substantial agreement for *tracing* movements, $\kappa = 0.705$ ($p < .001$), and also substantial agreement for *other* movements, $\kappa = 0.755$

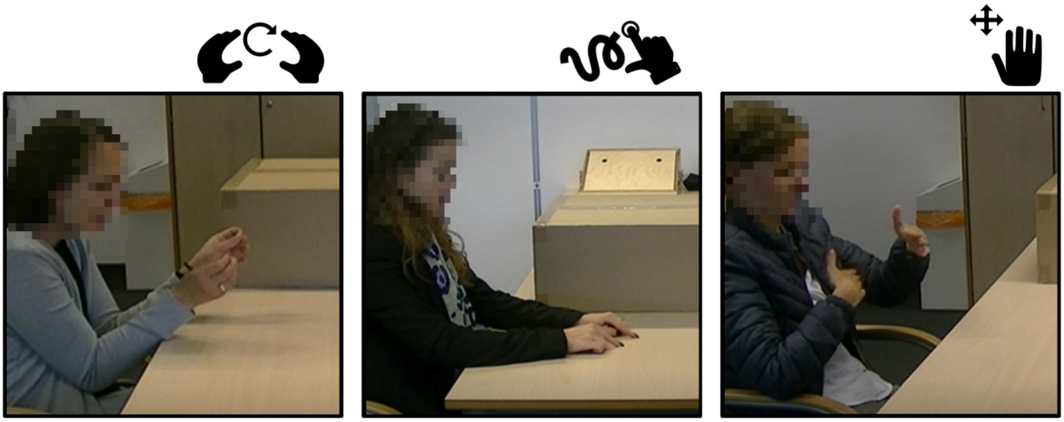

Fig. 2. Examples of type of gestures produced. From left to right: *Rotation* movement, *tracing* movement, and *other* movement.

($p < .001$).

1.6. Data exclusions

After the experiment, the experimenter asked participants who successfully reported a correct alternate interpretation whether they discovered the second interpretation during the perception phase or in the targeted imagery phase. All successful participants reconfirmed that they had not prematurely detected the alternate interpretation during perception. Those who did were already excluded in the perception phase as they would voluntarily report both interpretations.

Recall that participants were exposed to one trial in the touch-only condition and one trial in the vision-only condition. However, during the perception phase (when participants offered their initial interpretation of the figure after examining it through vision or touch), some participants offered an interpretation that suggested they were already aware of the bistability of the figure. Specifically, they gave both interpretations (the head interpretation and the body interpretation), or they gave an interpretation that matched the alternate orientation from what they had perceived (they were given the head orientation but interpreted the figure in the body orientation). Such responses make it difficult to know whether reinterpretation of the figure is occurring during imagery, or whether it has already occurred during perception (see Kamermans et al., 2017). Because we are specifically interested in participants' ability to reinterpret the figures in their mental imagery, we excluded trials where participants' responses during perception suggested bistability. This resulted in the exclusion of 26 trials (19 in the visual condition and 7 in the haptic condition). In addition, one participant's hands were not visible in the haptic condition making gesture coding impossible, and that trial was excluded from analysis as well.

1.7. Data analysis

For all analyses, we used the lme4 package in R (Bates, Sarkar, Bates, & Matrix, 2007) for logistic mixed regression models and nlme (Pinheiro et al., 2013) for linear mixed regression that included relevant fixed factors as well as random intercepts for each participant. To determine significance, we used the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) which applies Satterthwaite's approximation to degrees of freedom to determine the significance of individually fixed predictors. We also compared the fit of the model with the fixed factor of interest included to a base model that included only the random factor and any other fixed factors that were in the initial model. For dichotomous dependent variables, we used the glmer function to create mixed logistic regression models. For continuous dependent variables, we used the lmer function to create mixed linear regression models.

2. Results

2.1. Does success reinterpreting mental images depend on encoding modality?

We examined whether participants' ability to successfully reinterpret the figure depended on the modality in which they had learned the figure. We conducted a logistic regression model, with successful reinterpretation as the dependent variable, and condition (visual-only vs. touch-only) as a fixed factor. We also included participant as a random intercept. We found no significant effect of condition, $B = -0.73$ 95% CI $[-3.22, 1.38]$, $SE = 1.10$, $z = 0.66$, $p = .51$. The model that includes condition as a factor did not explain significantly more variance than the model that includes only the random intercept for participant, $X^2(1) = 0.46$, $p = .50$. Of the 49 participants who completed a vision-only trial, 10 (20%) offered a successful reinterpretation of the figure during the imagery task, compared to 16 of the 60 participants who completed a touch-only trial (26.7%). It appears that success at reinterpreting the figure using mental imagery is difficult (accomplished by only one-fourth of all participants) and does not depend on the modality in which the figure was learned.

2.2. Does perception modality affect the amount of gesture produced during mental imagery task?

First, we examined the likelihood that participants gestured in the vision-only versus touch-only trials at all. Of the 60 participants who completed the touch-only trial, 52% of them gestured at least one time during their explanation. Of the 49 participants who completed the vision-only trial, only 33% of them gestured at least one time during their explanation. To analyze this pattern statistically, we conducted a mixed logistic regression model with whether a gesture was produced as the dichotomous dependent variable. The model included condition (vision-only vs. touch-only) as the fixed factor and a random intercept for participant. The odds of a gesture occurring were 3.08 times higher in the touch-only condition than in the vision-only condition ($B = 1.126$, 95% CI $[0.22, 2.16]$, $SE = 0.479$, $z = 2.35$, $p = .02$). The model that includes condition explains significantly more variance in whether someone gestured or not than the model that includes only the random intercept for participant, $X^2(1) = 6.08$, $p = .010$.

Next, we considered the amount of time participants spent gesturing in each condition. The 31 participants who gestured at least once in the touch-only condition spent on average 13.55 s ($SD = 10.89$) gesturing, while the 16 participants who gestured at least once in the visual-only condition spent only 10.94 s ($SD = 6.71$) gesturing. We modeled this difference with a mixed regression that included condition (visual-only vs. touch-only), total amount of time spent in the imagery phase, and total number of words produced as fixed factors. A random intercept

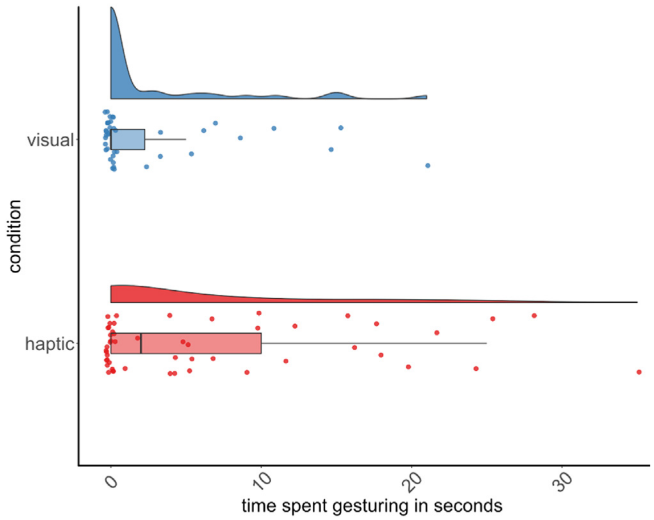

Fig. 3. The average time in seconds spent gesturing by participants as they described their reinterpretation of figures they had learned via vision-only or touch-only. The distributions reflect smoothed density distributions of the observations. The individual jitter dots reflect observed (time) gesturing per trial. The box plots reflect regular quartile intervals.

Table 1
Descriptives per condition.

Condition	Task time in seconds *M* (*SD*)	Time gestured in seconds *M* (*SD*)	Proportion gestured *M* (*SD*)	Number of words *M* (*SD*)
Vision-only	36.98 (27.42) (*n* = 48)	2.76 (5.81) (*n* = 58)	0.099 (0.208) (*n* = 49)	28.10 (23.48) (*n* = 48)
Touch-only	32 (26.14) (*n* = 62)	6.23 (10.09) (*n* = 66)	0.225 (0.316) (*n* = 60)	28.36 (23.42) (*n* = 61)

was included for participant. Not surprisingly, the time participants spent gesturing was positively related to the total amount of time they spent on the task ($B = 0.09$, 95% CI [0.009, 0.132], SE = 0.04, $t = 2.39$, $p = .03$) as well as to the number of words they spoke ($B = 0.17$, 95% CI [0.099, 0.236], SE = 0.04, $t = 4.41$, $p = .003$) in the imagery phase. However, there was also a significant effect of condition ($B = 6.76$, 95% CI [3.54, 7.37], SE = 1.49, $t = 4.52$, $p = .006$), such that participants who gestured spent significantly more time gesturing in the touch-only condition than in the vision-only condition (see Fig. 3). Including condition explains significantly more variance in the amount of time participants spent gesturing than the model that includes only time on task, the number of words spoken, and the random intercept for participants, $X^2(1) = 8.34$, $p = .003$. Descriptives for these measurements can be found in Table 1.

We next considered the specific form of gestures that were produced by those who gestured in each condition. The likelihood of producing a rotation gesture was not affected by condition. Nine (29%) of the 31 participants who gestured in the haptic condition produced at least one rotation gesture, while 3 (19%) of the 16 participants who gestured in the visual condition produced at least one rotation gesture. This difference was not statistically reliable ($B = -1.25$, $SE = 1.14$, $z = 1.10$, $p = .27$), and including condition as a factor does not explain more variance than excluding it, $X^2(1) = 1.51$, $p = .22$. There was also no difference between conditions in the likelihood of producing an “other” gesture, $B = 0.64$, $SE = 0.89$, $z = 0.72$, $p = .47$; $X^2(1) = 0.55$, $p = .45$. In contrast, the likelihood of producing a tracing gesture was affected by condition. Twelve of the 31 (39%) participants who gestured in the touch-only condition produced at least one gesture that seemed to trace the contours of the figure, while 0 participants did this in the vision-only condition. Modelling this difference is problematic due to the lack of variability in the vision-only condition, but the pattern is clear.

We also considered whether condition predicted time spent gesturing for a particular type of gesture. We performed a mixed regression with nlme (random intercept participant; random slope did not converge), to assess whether time gesturing was predicted by gesture type (tracing, rotation, other), condition (visual, haptic) and their interaction. Compared to a model predicting the overall mean, a model containing gesture type reliably improved predictions of time, $X^2(1) = 19.24$, $p < .001$. Furthermore, adding condition to this previous model further improved predictions, $X^2(1) = 5.25$, $p = .022$. Adding an interaction between gesture type and condition did not improve predictions of time further, $X^2(1) = 1.85$, $p = .395$. The best model containing the main effects of condition and gesture type revealed that less time was spent gesturing in the vision-only condition than in the touch-only condition, $B = -1.14$, $SE = 0.50$, $t(264) = 2.28$, $p = .023$. Furthermore, Bonferroni corrected post-hoc testing (using lsmeans) revealed that there was no difference in the amount of time spent producing rotation versus trace gestures, diff = −0.35, SE = 0.60, $p = .99$, but less time was spent producing Other types of gestures, as compared to the trace as well as the rotate gestures (diff's < 2.1, p's < .001).

It appears that, as predicted by the GSA framework, having manual experience during the perception of an image as opposed to only visual experience, leads to significantly more gestures when the image is thought about, manipulated, and described. Further, participants are particularly likely to gesture about the shape of the object (tracing its contours) after having experience manually exploring it during encoding.

2.3. Is gesture in the imagery task related to reinterpretation success?

To address this question, we modeled whether participants offered a correct reinterpretation of the figure during the imagery phase in a mixed logistic regression that included condition, whether participants ever gestured, and their interaction as fixed factors. A random intercept was included for participant. There were no significant effects observed, and the model containing the fixed factors did not explain more variance in reinterpretation success than the model containing only the random intercept for participant, $X^2(3) = 6.17$, $p = .10$. Thus, we see no evidence in this study that gestures are associated with more success on the reinterpretation task.

3. Intermediate discussion

In Experiment 1, we see evidence that participants were more likely to gesture about a figure they had learned through manual exploration than about a figure they had learned through vision. They were particularly likely to produce gestures that traced the contours of the shape they had learned, which presumably mimic the actions that were produced during manual exploration (see Lederman & Klatzky, 1993). This is in line with previous reports that speakers gesture more about images they have physical experience interacting with than images they have only seen (Hostetter & Alibali, 2010) and supports the central claim of the GSA framework (Hostetter & Alibali, 2008) that gestures emerge when speakers reactivate action experiences that were performed during perception. When these experiences were manual (in the touch-only condition), more gestures occurred as the perceptual experience was re-evoked in imagery than when these experiences were not manual (in the vision-only condition).

In contrast, no significant difference was found between conditions for rotation gestures. According to the GSA framework, rotation gestures occur as a result of imagining the movement and action involved to complete a transformation (see Hostetter & Alibali, 2018). In the present paradigm, this imagined movement could happen regardless of whether the figure was learned through touch or through vision, and indeed, the central task participants were given in both conditions was to engage in this type of imagined movement. Even though participants had experience touching the figures in the touch-only condition, they did not have experience physically rotating the figures and there was no reason to avoid imagining manual rotation of the figures in the vision-only condition. Rather, we contend that manual rotation could be similarly imagined in both conditions, and was equally likely to be expressed in gesture in both conditions.

We should further note that despite similar performance between haptic and visual conditions in imagery performance, the data exclusions show clear differences in the way haptic versus visual detection of bistability occurs in perception. Specifically, participants in the visual condition were likely to already detect bistability during perception, while this was less likely in the haptic condition. A possible explanation is that visual processes occur on faster timescales wherein saccades are rapidly detecting shape segments of an object, while haptic processes occur on slightly slower timescales as the fingers trace the object. Yet, despite clear differences in visual versus haptic perception of bistability, detection of bistability in mental imagery did not seem to rely on those differences as we did not find performance differences in mental imagery. It is in this interesting way that perception and imagery does not behave equivalently.

Regarding gestures, we see no evidence that the gestures produced by participants after manual exploration were beneficial to the reinterpretation task. That is, participants were not more likely to successfully reinterpret their image if they gestured than if they did not gesture. However, the majority of gestures produced by participants occurred with speech. We observed very few gestures occurring before participants began giving their answer in the imagery phase; only three participants produced such gestures. So-called co-thought gestures (Chu & Kita, 2011) may be more likely to facilitate thinking than co-speech gestures, which may be produced after speakers have already formed their best guess for a reinterpretation of the image. Perhaps gestures would be more helpful for participants in this task if they were indeed produced as participants were more actively engaged in thinking about the image (rather than in describing their reinterpretation). To investigate this possibility, we conducted Experiment 2, in which some participants were instructed to use their hands as they attempted to reinterpret their image. If producing gestures benefits participants' ability to reinterpret their mental images, then participants who are instructed to gesture should have better performance than those who do not gesture.

In Experiment 2, all participants learned the figure through touch (rather than through vision). We decided to drop the vision-only condition for two reasons. First, recall that our hypothesis about the functionality of gesture for visual imagery hinges on the possibility that activating bodily movements involved in perception might strengthen information about a figure's shape. As such, we are particularly interested in gestures that resemble the actions involved in haptic exploration in the touch-only condition. Although manual gestures could resemble actions involved in visual perception, the connection is less straightforward. Second, in Experiment 1, we found that gestures were most prevalent in the touch-only condition. Thus, it seems most prudent to examine the potential functionality of gestures in a condition where they are particularly likely to occur spontaneously. Our hypothesis is that using gesture to recreate the actions involved in perception might make it easier to find the alternative interpretation of the figures. An alternative possibility is that producing such gestures is not critically important, but the ability to freely imagine interacting with the figures (even without externalizing the movements as gesture) is important for reinterpretation success. For example, Klatzky, Lederman, and Matula (1991) found that participants consistently report imagining active manual exploration of objects when asked to judge properties of the object such as “roughness” or “hardness”. Moreover, Eardley and Pring (2007) demonstrated that both early blind and blindfolded-sighted participants performed significantly worse on a mental synthesis task when experiencing spatial interference. Thus, interfering with participants' ability to imagine manual exploration could significantly interfere with their ability to perform the reinterpretation task. To test this claim, we required some participants in Experiment 2 to engage in manual tapping as they attempted to reinterpret their image, in order to interfere with their ability to imagine meaningful manipulation of the figure. This is similar to the manipulation used in previous studies that have sought to prevent participants from engaging in premotor planning as they engage with a task (e.g., Brooks, Barner, Frank, & Goldin-Meadow, 2018; Frank & Barner, 2012). If imagined interaction with the figure is important for reinterpretation success, performance in the motor interference condition should be worse than in either the no gesture or gesture conditions. Note that these two hypotheses are not mutually exclusive; imagined movement could be beneficial for reinterpretation success, and externalizing that imagined movement in gesture could provide a further boon.

Finally, if imagined movement or gesture is beneficial for reinterpretation, they might be particularly beneficial for those who have fewer internal resources at their disposal to solve the task. For example, Eielts et al. (2018) found that gestures were particularly beneficial for solving the Tower of Hanoi task for participants who had low visual working memory capacity (also see Pouw et al., 2016). Therefore, in order to assess whether gesture is particularly beneficial for mental imagery for those participants who have fewer internal resources to employ for the task, we measured each participant's visual working memory capacity.

4. Method Experiment 2

4.1. Participants and design

All participants explored one of the bistable figures through touch-only. Then, participants were randomly assigned to one of three between-subjects conditions as they attempted to find the alternative interpretation of the figure; instructed gesture, instructed no manual movement, and manual interference. The main dependent variable was the correct interpretation (no [correct] interpretation vs. correct interpretation; similar to Experiment 1). We opted for a between-subjects design as research has shown that potential effects of gestures on mental imagery (i.e., mental rotation task) may carry over to subsequent trials without movement (Chu & Kita, 2011). Additionally, it prevents high exclusion rates, as it lowers the chance of premature bistability detection (i.e., the chance of recognizing that figures are ambiguous increases when seeing more figures).

We recruited 126 students who participated for course credits or for a small monetary reward (2.50 euro). Based on power calculations performed with G*Power software (version 3.0.10), a sample size of 108 participants ($N = 36$ per condition) is needed in order to detect a difference of medium effect size between conditions for a dichotomous outcome (estimated for a Chi-Squared Test) with 80% Power, alpha = 0.05, medium effect size ($w = 0.3$), $df = 2$. As such, we continued the experiment until we had at least 108 participants with an approximately balanced number of trials for the different orientations and conditions. The final dataset after exclusions, includes data from 114 participants (Mage = 21.10 years, $SD = 3.21$, 20% Male; Native Dutch students = 22.08%, International students = 77.02%). Although some participants ($N = 23$) did not fill out their age on the questionnaire, all participants were students.

4.2. Materials

4.2.1. Test figures

We used one bistable figure for all participants, namely the deer/seal. We chose the deer/seal figure because it produced the lowest rate of premature bistability detection during memorization in Experiment 1 in comparison to the penguin/giraffe figure.

4.2.2. Visual working memory task

To assess visual working memory capacity (VWM), we administered a modified version of the Visual Patterns Test (Della Sala, Gray, Baddeley, & Wilson, 1997) programmed in Adobe Flash (as used in Chu et al., 2014a, 2014b, Eielts et al., 2018; Pouw et al., 2016).[2] The task consisted of 25 trials, preceded by two practice trials. Each trial consisted of a 3-s presentation of a visual pattern consisting of black rectangular cells (14 cm × 14 cm) presented in a matrix where half of the cells were empty. Following the initial presentation, the same matrix appeared with non-filled cells. Participants recreated the pattern of black squares by selecting-and-clicking the cells that were previously filled (i.e., colored black). If participants (failed to) completely recreate the pattern the trial was scored as (in)correct. The maximum score was 25 points (1 per correct trial), with higher scores suggesting higher visual working memory capacity.

4.2.3. Questionnaire

Similar to Experiment 1, participants reported their Age, Sex, and Native Language, and their perceptions regarding the nature of the experiment and the expectations of the researchers. Importantly, we added another exploratory question to tap into their conscious experience during haptic imagery. That is, we asked whether they could "explain as detailed as possible what you experienced during the mental imagery, for example: did you imagine *feeling and touching* or imagine *seeing* the figure when trying to detect another interpretation?". To provide a summary descriptive, we coded each answer as visual, haptic, or multimodal (i.e., visual and haptic) (see Table 2).

Table 2
Self-reported perceptual experience during mental imagery.

Experience	Count
Visual imagery experience	53/114 (46.50%)
Haptic imagery experience	11/114 (9.65%)
Visual-haptic imagery experience	35/114 (30.70%)
Not-determined	15/114 (13.15%)

4.3. Procedure

Participants, none of whom participated in Experiment 1, were informed that the research was about imagination. Participants were video-recorded for the total duration of the experiment.

4.3.1. Perception phase

The perception phase was similar to the perception phase in the touch-only condition of Experiment 1. Participants inspected the deer/seal figure for 30 s with touch only, with half inspecting the figure in the head (deer) orientation and half inspecting the figure in the body (seal) orientation. Participants then reported what they thought the shape was.

4.3.2. Imagery phase

The imagery phase was similar to Experiment 1, wherein participants were instructed to close their eyes and bring back the memorized figure in mental imagery. They were then told about bistability and asked to reinterpret their image by mentally rotating the figure. Three additional instructions were given during the imagery phase depending on condition.

In the gesture condition, participants were asked to: "try physically moving your hands around the contours of the figure. Thus try to move the hands as-if actually having the figure in your hands. Try to do this continuously to bring a clear memory back of the figure". We chose to emphasize *contour following* as this type of gesture was observed in Experiment 1 and has been identified as one of the stereotyped manual movements that are made to identify shape (Klatzky, Lederman, & Matula, 1993). Participants were then informed about ambiguity of the figure and were asked to rotate their mental image 180 degrees upside down to find another interpretation with the instruction to move their hands as if actually rotating the figure. If participants were not moving their hands, they were prompted to try to imagine actually feeling the figure.

In the manual interference condition, participants were asked to close their eyes and gently drum their fingers with both hands continuously on the table at their own preferred pace during the imagery phase (a similar procedure has been used by Frank & Barner, 2012). This manual tapping was continued as participants reimagined the figure as well as during their attempts to reinterpret it.

Finally, in the no gesture condition, participants were instructed to keep their hands flat on the table through the duration of the imagery phase. This was done as to ensure that participants would not spontaneously gesture as they completed the task (as observed in Experiment 1).

4.4. Performance scoring

4.4.1. Reinterpretation

As in Experiment 1 (and Kamermans et al., 2017), participants were primarily their own raters; i.e., we counted reinterpretations during imagery as correct if they were given by another participant during the perception phase.

4.4.2. Exclusions

Twelve participants were excluded as they a) reported during perception an interpretation that belonged to the opposite orientation ($N = 10$), b) reported that they misunderstood the instructions ($N = 1$), or c) were given incorrect instructions by the experimenter ($N = 1$). This resulted in a final sample of 114 participants, divided across the gesture ($n = 38$), no gesture ($n = 35$), and manual interference ($n = 41$) conditions.

4.4.3. Descriptives self-report

Table 2 shows an overview of the proportion of participants reporting mental imagery experience as being predominantly visual, predominantly haptic, or multimodal (visual and haptic). Although all participants encoded the figure through touch, the majority (77.20%) reported having primarily visual or multimodal (visual and haptic imagery) experiences. However, note that 40.35% also reported having experienced at least some form of haptic imagery. There were fifteen cases where it was not clear from the answers whether participants experienced visual, haptic, or multimodal images.

5. Results Experiment 2

5.1. Does reinterpretation success depend on manual activity?

Table 3 shows the proportion of successful reinterpretations per condition and retrieval target. To assess the effect of condition on reinterpretation performance, we performed logistic regression in R. Participants in the manual interference condition had significantly worse reinterpretation performance than those in the gesture condition, $B = -1.445$ 95% CI $[-2.592, -0.407]$, SE = 0.551, $z = -2.651$, $p = .009$, as well as those in the no gesture condition, $B = -1.311$ 95% CI $[-2.471, -0.252]$, SE = 0.558, $z = -2.348$, $p = .019$. Furthermore, reinterpretation success in the gesture condition did not reliably differ from reinterpretation success in the no gesture condition, $B = 0.134$ 95% CI $[-0.797, 1.0711]$, SE = 0.474, $z = 0.282$, $p = .778$.

---

[2] This task can be retrieved from https://osf.io/725te/.

Table 3
Successful reinterpretation rate.

	No gesture	Gesture	Manual interference	Total
Deer	9/18 (50.00%)	9/19 (47.37%)	6/20 (30.00%)	24/57 (42.11%)
Seal	5/18 (27.78%)	6/19 (31.58%)	1/21 (4.77%)	12/58 (20.69%)
Total	14/36 (40%)	15/38 (39.47%)	7/41 (17.07%)	36/115 (31.58%)

Note: There were some participants who could not produce an interpretation during the perception phase ($N = 13$) but were nevertheless included as we were interested in the discovery of a novel interpretation in *imagery* of the rotated figure. Seven of these thirteen participants produced an interpretation in imagery (correct interpretations = 4).

5.1.1. *Visual working memory capacity*

Mean score on the visual working memory task (VWM) was $M = 0.758$, $SD = 0.156$. No sampling bias of condition assignment was observed; the imagery condition ($M = 0.77$, $SD = 0.162$, 95% CI [0.71, 0.82]), enactment condition ($M = 0.78$, $SD = 0.162$, 95% CI [0.72, 0.84]), and interference condition ($M = 0.73$, $SD = 0.16$, 95% CI [0.68, 0.79]) did not show significant differences on VWM ($F(2, 102) = 0.303$, $p = .739$).

Further, there were no indications that VWM was related to interpretation performance, in any of the conditions. There was no significant correlation between retrieval time for correct interpretations and VWM, $r = -0.0176$, $p = .858$ (nor did we find significant correlations within conditions). Additionally, through a logistic regression, we obtained that VWM did not reliably predict successful interpretations after accounting for variance explained by condition, $b = 1.513$, $SE = 1.40$, $z = 0.108$, $p = .914$, and no reliable interactions of condition and VWM were obtained.

6. General discussion

In Experiment 1, we observed that participants were more likely to spontaneously gesture during the task when they had previously felt a figure compared to when they had only seen that figure. However, participants who gestured were no more successful at reinterpreting their mental image than those who did not gesture. Most gestures occurred with speech (i.e., when participants described their answer) rather than no speech (i.e., during the act of mentally reinspecting the figure). Experiment 2 explored if co-thought gestures would be beneficial to performance by encouraging some participants to gesture during the mental imagery task itself. Results of this experiment showed that participants who were instructed to gesture did not reliably outperform those who kept their hands still. However, both groups (those who gestured and those who kept their hands still) outperformed participants who engaged in a manual interference task.

Our finding that gestures occur more frequently following manual as opposed to visual exploration of a figure is in line with the GSA framework (Hostetter & Alibali, 2008, 2018). Namely, the GSA framework suggests that experiences involved during perception are reactivated during imagery, and that these simulations are particularly likely to be expressed as gestures when they involve manual action. This finding adds to previous reports that gestures are more prevalent when speakers describe objects that they have interacted with (Hostetter & Alibali, 2010) or that they can readily imagine interacting with (Chu & Kita, 2016) than when they describe information that is less closely tied to action. It appears that gestures can reflect imagined interactions with objects.

Further, we found suggestive evidence that *imagined* manual interaction with the figures was functionally involved in the reinterpretation task. Specifically, in Experiment 2, participants who were engaged in manual interference (i.e., tapping) as they attempted to reinterpret the figures were less successful than those who were able to freely imagine manually interacting with the figures. This negative effect of motor interference has been demonstrated in other imagery tasks as well. For example, Brooks et al. (2018) found that motor interference (tapping on a keyboard) negatively affected children's performance on a mental abacus task (also see Frank & Barner, 2012), whereas production of meaningful gestures did not reliably affect performance. Hegarty et al. (2005) found that spatial tapping interfered with participants' ability to successfully solve mental animation problems which involved imagining how components of a diagram would move in relation to one another (also see Eardley & Pring, 2007). Nathan and Martinez (2015) found that tapping interfered with participants' ability to make spatial inferences about the function of the circulatory system. Similarly, we found that preventing participants from imagining how they would interact with the figures (by requiring them to use their hands for an irrelevant tapping task) significantly impaired their ability to interpret the result of mentally rotating the figure. It appears that the ability to freely imagine motor movement is critically important for being able to judge the spatial and visual consequences of that movement.

In our task, motor system interference could have prevented participants from performing the spatial rotation necessary to reorient the figure or it could have prevented participants from seeing the visual result of that rotation in their imagery (or both). A disruption in either process would result in the impaired performance we observed on the reinterpretation task. However, given existing evidence about the importance of the motor system for spatial processing (e.g., Postle et al., 2006), we suspect that motor interference likely made it difficult for participants to imagine the spatial transformation of the figure, which of course also then made it difficult for them to generate a correct reinterpretation because their image did not reach the correct endstate. Whether motor interference further affected participants' ability to *visually* interpret the endstate of that rotation (in situations where it was completed successfully) is unclear from these data.

Although an inability to freely imagine movement impaired performance on this task, we found no evidence that enacting movements in gesture benefited performance. Yet gestural movement (as a concurrent task) is not detrimental for imagery performance, suggesting that it is not simply any kind of movement that interferes with task performance.

Although gesture does not interfere with task performance, in Experiment 1, participants who spontaneously produced gestures as they explained their reinterpretation were *no more successful* than those who did not. Further, in Experiment 2, participants who were instructed to gesture as they thought about the figure were no more likely to devise a correct reinterpretation than those who kept their hands flat on the table. This is in contrast to previous reports that gestures benefit performance in other types of imagery tasks, such as mental rotation (Chu & Kita, 2011), mental abacus calculations (Cho & So, 2018), and mental problem solving of the tower of Hanoi (Eielts et al., 2018).

We speculate that a key difference between those tasks and this reinterpretation task is the differential importance of *shape-specific* versus *spatial-temporal* visual imagery to the tasks. In tasks like mental rotation or Tower of Hanoi, it is primarily being able to track the spatial locations of the components in the image through time (the discs in tower of Hanoi, the beads in the abacus, the block segments in mental rotation) that determines success on the task. In contrast, in the reinterpretation task used here, participants had to be able to “see” the endstate of the figure's rotation in order to correctly identify it. It is possible that gestures are particularly helpful for generating or maintaining spatial-temporal information about relative location, while being less helpful for generating or maintaining visual(−haptic) information about object shape. Note, however, that this explanation is at apparent odds with previous reports which show specifically *shape-specific visual* (rather than spatial-temporal) working memory capacity that is most predictive of the occurrence (Chu et al., 2014a, 2014b) and benefit (Eielts et al., 2018; Pouw et al., 2016) of gesture.

Another possibility is that participants were not experienced enough with our task for gestures to be helpful. For example, Cho and So (2018) found that gestures were only helpful on a mental abacus task for participants who had an intermediate amount of experience with the task. When participants had a lot of experience with the mental abacus task, they appeared able to mentally imagine the movements necessary without externally producing them (see also Brooks et al., 2018). Further, when participants were just beginning to learn the task, gestures were also not helpful, likely because they did not yet fully understand the correspondence between the motor act of adjusting the abacus and its spatial consequences. Future research should investigate if participants with more experience perceiving the test figures through touch alone would come to benefit from gestures that enact their imagery of the figure's shape.

In addition to their relevance for understanding gesture, our data also have implications for understanding the involvement of the motor system in imagery of haptic experiences. After learning the figure through touch in Experiment 2, about 40% of participants explicitly reported a haptic experience of imagining feeling the figure during imagery or a mixed haptic and visual experience (see Table 2). This suggests that imagery of a haptic experience often involves the subjective experience of kinesthetic haptic activation. Further, in Experiment 1, 39% of participants who had learned the figure through touch externalized imagined tracing movements in their gesture as they described the image. Thus it seems that reactivation of kinesthetic experiences is important for imagery of a haptic experience, just as eye movements have been shown to be important for imagery of a visual experience (e.g., Laeng et al., 2014).

At the same time, it is also worth noting that 77% of participants reported experiencing visual imagery (or a combination of visual and haptic imagery) after learning the figures through touch in Experiment 2. This suggests that, at least in this task, many participants were attempting to imagine not just how the figure felt but also how it looked. Indeed, it may be that the task of detecting bistability in these figures requires the creation of a holistic understanding of the entire figure - an understanding which is more easily accomplished with the fast saccades of the visual modality than with the relatively slow finger traces of the haptic modality. For example, we found in Experiment 1 that participants often detected bistability during visual perception, while this happened less frequently during haptic perception. In future work, it may be interesting to devise ways of manipulating not just the experience participants have as they memorize a figure, but the type of imagery (haptic versus visual) they evoke as they think about it.

Although the current paradigm of reinterpretation in mental imagery has been widely adopted (e.g., Chambers & Reisberg, 1985; Kamermans et al., 2017; Mast & Kosslyn, 2002; Peterson et al., 1992), there are some inherent shortcomings that should be addressed in future research. Most notably, scoring correct versus incorrect interpretations may be to some extent arbitrary. We have tried to counteract this by using participants as their own raters (similar to Kamermans et al., 2017), thereby minimizing post-hoc decisions that had to be made. As such, our procedure can be seen as a more conservative and reliable procedure than solely relying on post-hoc experimenter ratings (cf. Chambers & Reisberg, 1985; Peterson et al., 1992). Nevertheless, this study is not immune to the possibility that misses and/or false positives in reinterpretation performance occurred. Future research should, therefore, focus on additional ways to objectify the qualitative content that is targeted by imagery. Another issue with the current paradigm is that one must control for the possibility that participants prematurely perceived the alternate interpretation during visual memorization. This often leads to a high number of exclusions, as was the case in Experiment 1 (see also Kamermans et al., 2017; Mast & Kosslyn, 2002). We have tried to counteract this in Experiment 2 by opting for a between-subjects design and using the stimulus figure that resulted in the lowest number of premature bistability detections in Experiment 1.

Finally, an important result that converges with research on reinterpretation in imagery is that it is relatively hard to do (Chambers & Reisberg, 1985; Kamermans et al., 2017; Mast & Kosslyn, 2002; Peterson et al., 1992). Across the board, no more than half of the participants are able to perform successful reinterpretations in visual imagery. Given the present data, this seems to be the case when the figure has been learned via touch as well (about 20–30%). Interestingly, it has been found that when participants are allowed to draw their visual memory of an ambiguous figure on paper, correct reinterpretations of the drawings are readily made (Chambers & Reisberg, 1985). It has been argued that this difference in performance is present because perception is differently constrained than imagination. It is not the quality of the “mental representation” per se that is problematic (because one can draw it out), but fundamental differences in processes of perception versus imagery that produce differences in performance. The present results suggest that “gesturing it out” is not as stable a perceptual platform as drawing it out. Yet at the same time the results also suggest that perception-action processes are still functionally involved in imagery, muddling the strict functional boundary between perception and imagery.

In conclusion, we have shown that reinterpreting a mental image of a figure that has been learned through touch involves imagined manual interaction with the figure, and these imagined movements are expressed alongside speech as gesture. Further, preventing participants from engaging in such imagined movement negatively affects their ability to reinterpret the image, even though preventing movement in gesture does not. These results support the idea that imagined movement is critically important for success at retrieving and mentally manipulating a figure that was learned haptically, and that gestures reflect these imagined movements even as they do little to further strengthen that imagery in this task.

Contributions

KK, WP, and AH were main contributors to writing the current manuscript, with critical revisions by FP, as well as LF and AA. WP developed the research idea with critical guidance by FP, and the design was majorly influenced by earlier research from KK, WP and FP. KK was main contributor to gesture coding analyses, with LF recoding part of the data for Experiment 1 and interrater reliability. WP coordinated data collection. AA collected the data for Experiment 1. LF collected data for Experiment 2. WP performed analyses for experiment 2. AH performed analyses for Experiment 1 and coordinated gesture coding procedures.

Declaration of Competing Interest

The data of Experiment 1 have been reported in the *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Acknowledgements

The authors would like to express their gratitude to Charly Eielts for programming the working memory task. This research was co-funded by NWO 446-16-012 (recipient: Wim Pouw). We would like to thank Jim Davies for fruitful discussions about this line of research.

Data access

Data and analyses scripts (not including the video data) supporting this research report can be retrieved from the Open Science Framework (https://osf.io/725te/).

References

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). *The lme4 package. R package version, 2(1).* 74.

Boncoddo, R., Dixon, J. A., & Kelly, E. (2010). The emergence of a novel representation from action: Evidence from preschoolers. *Developmental Science, 13*, 370–377. https://doi.org/10.1111/j.1467-7687.2009.00905.x.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*, 27–38. https://doi.org/10.1162/jocn.1997.9.1.27.

Brooks, N. B., Barner, D., Frank, M., & Goldin-Meadow, S. (2018). The role of gesture in supporting mental representations: The case of mental abacus arithmetic. *Cognitive Science, 42*(2), 554–575.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance, 11*(3), 317–328.

Cho, P., & So, W. C. (2018). A feel for numbers: The changing role of gesture in manipulating the mental representation of an abacus among children at different skill levels. *Frontiers in Psychology, 9*, 1267.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*, 706–723. https://doi.org/10.1037/a0013157.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140*, 102–116. https://doi.org/10.1037/a0021790.

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: General, 140*, 102–116. https://doi.org/10.1037/a0021790.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014a). 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.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014b). 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.

Della Sala, S., Gray, C., Baddeley, A., & Wilson, L. (1997). *The Visual Patterns Test: A new test of short-term visual recall.* Suffolk: Thames Valley Test Company.

Eardley, A. F., & Pring, L. (2007). Spatial processing, mental imagery, and creativity in individuals with and without sight. *European Journal of Cognitive Psychology, 19*(1), 37–58.

Ehrlich, S. B., Levine, S., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology, 2006*(42), 1259–1268.

Eielts, C., Pouw, W., Ouwehand, K., Van Gog, T., Zwaan, R. A., & Paas, F. (2018). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychological Research*. https://doi.org/10.1007/s00426-018-1065-9.

Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science, 13*, 51–78. https://doi.org/10.1016/0364-0213(89)90011-6.

Foglia, L., & O'Regan, K. J. (2016). A new imagery debate: Enactive and sensorimotor accounts. *Review of Philosophy and Psychology, 7*, 181–196. https://doi.org/10.1007/s13164-015-0269-9.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General, 141*(1), 134.

Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science, 15*, 876–884. https://doi.org/10.1111/j.1467-7687.2012.01185.x.

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation, 5*, 333–356. https://doi.org/10.1207/s15427633scc0504_3.

Hostetter, A., Alibali, M., & Bartholomew, A. (2011, January). Gesture during mental rotation. *Proceedings of the annual meeting of the Cognitive Science Society. Vol. 33* (No. 33).

Hostetter, A. B. (2014). Action attenuates the effect of visibility on gesture rates. *Cognitive Science, 38*(7), 1468–1481.

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. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language, 63*, 245–257. https://doi.org/10.1016/j.jml.2010.04.003.

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.

Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science, 25*, 236–242. https://doi.org/10.1177/0956797613498260.

Kamermans, K. L., Pouw, W., Mast, F. W., & Paas, F. (2017). Reinterpretation in visual imagery is possible without visual cues: A validation of previous research. *Psychological Research,* 1–14.

Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1991). Imagined haptic exploration in judgments of object properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 314–322. https://doi.org/10.1037/0278-7393.17.2.314.

Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1993). Haptic exploration in the presence of vision. *Journal of Experimental Psychology: Human Perception and Performance, 19*(4), 726.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery.* Oxford University Press.

Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition, 33*(4), 710–726.

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

Laeng, B., Bloem, I. M., D'Ascenzo, S., & Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. *Cognition, 131*, 263–283. https://doi.org/10.1016/j.cognition.2014.01.003.

Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science, 26*, 207–231. https://doi.org/10.1016/S0364-0213(01)00065-9.

Lederman, S. J., & Klatzky, R. L. (1993). Extracting object properties through haptic exploration. *Acta Psychologica, 84*, 29–40.

Leo Burnett Worldwide (2015). See whatever you want to see [Advertisement]. Retrieved from https://www.brandingmag.com/2013/12/19/jeep-see-whatever-want-see/.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479–1498.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience, 31*(3), 430–440.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2017). How what we see and what we know influence iconic gesture production. *Journal of Nonverbal Behavior, 41*(4), 367–394.

Mast, F. W., & Kosslyn, S. M. (2002). Visual mental images can be ambiguous: Insights from individual differences in spatial transformation abilities. *Cognition, 86*(1), 57–70.

Nathan, M. J., & Martinez, C. V. J. (2015). Gesture as model enactment: The role of gesture in mental model construction and inference making when learning from text. *Learning: Research and Practice, 1*, 4–37. https://doi.org/10.1080/23735082.2015.1006758.

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.

Oliver, R. T., & Thompson-Schill, S. L. (2003). Dorsal stream activation during retrieval of object size and shape. *Cognitive, Affective, & Behavioral Neuroscience, 3*(4), 309–322.

Pearson, D. G., & Logie, R. H. (2015). A sketch is not enough: Dynamic external support increases creative insight on a guided synthesis task. *Thinking & Reasoning, 21*(1), 97–112.

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences, 112*(33), 10089–10092.

Peterson, M. A., Kihlstrom, J. F., Rose, P. M., & Glisky, M. L. (1992). Mental images can be ambiguous: Reconstruals and reference-frame reversals. *Memory & Cognition, 20*(2), 107–123.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H. C., Cohen, L., ... Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences, 101*(15), 5658–5663.

Pine, K. J., Gurney, D. J., & Fletcher, B. (2010). The semantic specificity hypothesis: When gestures do not depend upon the presence of a listener. *Journal of Nonverbal Behavior, 34*, 169–178. https://doi.org/10.1007/s10919-010-0089-7.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C (2013). nlme: Linear and nonlinear mixed effects models. *R package version, 3(1)* (pp. 111). .

Postle, B. R., Idzikowski, C., Sala, S. D., Logie, R. H., & Baddeley, A. D. (2006). The selective disruption of spatial working memory by eye movements. *The Quarterly Journal of Experimental Psychology, 59*(1), 100–120.

Pouw, W., & Hostetter, A. (2016). Gesture as predictive action. *Reti, Saperi, Linguaggi: Italian Journal of Cognitive Sciences, 3*, 57–80. https://doi.org/10.12832/83918.

Pouw, W. T. J. L., de Nooijer, J. A., van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology, 5*, 359. https://doi.org/10.3389/fpsyg.2014.00359.

Pouw, W. T. J. L., Mavilidi, M., 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, 17*, 269–277. https://doi.org/10.1007/s10339-016-0757-6.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*(2), 157–182. https://doi.org/10.1017/S0140525X02000043.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703.

Sima, J. F., Schultheis, H., & Barkowsky, T. (2013). Differences between spatial and visual mental representations. *Frontiers in Psychology, 4*, 240.

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex, 15*(10), 1570–1583.

Smithson, L., & Nicoladis, E. (2014). Lending a hand to imagery? The impact of visuospatial working memory interference upon iconic gesture production in a narrative task. *Journal of Nonverbal Behavior, 38*(2), 247–258.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research, 65*(4), 235–241.

Spivey, M. J., Tyler, M. J., Richardson, D. C., & Young, E. E. (2000, January). Eye movements during comprehension of spoken descriptions. *Proceedings of the Annual Meeting of the Cognitive Science Society*. *Vol. 22* (No. 22).

Stieff, M., Lira, M. E., & Scopelitis, S. A. (2016). Gesture supports spatial thinking in STEM. *Cognition and Instruction, 34*, 80–99. https://doi.org/10.1080/07370008.2016.1145122.

Thomas, N. J. (1999). Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science, 23*(2), 207–245.

Wallraven, C., Bülthoff, H. H., Waterkamp, S., van Dam, L., & Gaißert, N. (2014). The eyes grasp, the hands see: Metric category knowledge transfers between vision and touch. *Psychonomic Bulletin & Review, 21*(4), 976–985.

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