

Capacity for Visual Features in Mental Rotation

Yangqing Xu and Steven L. Franconeri

Northwestern University

Psychological Science
2015, Vol. 26(8) 1241–1251
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797615585002
pss.sagepub.com


Abstract

Although mental rotation is a core component of scientific reasoning, little is known about its underlying mechanisms. For instance, how much visual information can someone rotate at once? We asked participants to rotate a simple multipart shape, requiring them to maintain attachments between features and moving parts. The capacity of this aspect of mental rotation was strikingly low: Only one feature could remain attached to one part. Behavioral and eye-tracking data showed that this single feature remained “glued” via a singular focus of attention, typically on the object’s top. We argue that the architecture of the human visual system is not suited for keeping multiple features attached to multiple parts during mental rotation. Such measurement of capacity limits may prove to be a critical step in dissecting the suite of visuospatial tools involved in mental rotation, leading to insights for improvement of pedagogy in science-education contexts.

Keywords

mental rotation, capacity, attention, selection, tracking, eye tracking, visual working memory

Received 6/16/14; Revision accepted 4/10/15

Creating and transforming representations of structure is a core component in almost all forms of reasoning (Hegarty, 2004; Stieff, 2007), and spatial thinking skills are central to success in science, technology, engineering, and mathematics (STEM) disciplines (Newcombe, 2010; Wai, Lubinski, & Benbow, 2009). Despite the importance of these abilities, students of both science and medicine have difficulty performing mental rotation (Hegarty, Keesner, Cohen, Montello, & Lippa, 2007; Stieff, Dixon, Ryu, Kumi, & Hegarty, 2014). Given the interest in training spatial thinking skills (National Research Council, 2006), researchers have sought strategies for improving them (e.g., Terlecki & Newcombe, 2005; for a review, see Uttal et al., 2013). But potential advances can be constrained by the fact that researchers do not yet fully understand the suite of underlying visuospatial tools that allow observers to construct and transform mental representations of structure.

One way to dissect this suite of tools is to dissociate the capacity limitations of potentially separable types of processing. Many types of structural representations might contribute to mental rotation. Such representations include not only representations of visual information,

but also motoric or other embodied representations (Zacks, 2008), verbal coding (Stieff & Raje, 2010), and long-term memories (Steiger & Yuille, 1983). Explorations of rotation capacity for visual information have focused almost exclusively on limitations on transforming shape-envelope information, including the structure of concatenated 3-D blocks (Shepard & Metzler, 1971; Yuille & Steiger, 1982), 2-D geometric drawings (Pylyshyn, 1979), and 2-D random polygons (Cooper & Podgorny, 1976; Folk & Luce, 1987). Some of the results obtained suggest a virtually unlimited capacity for shape transformation, such that a detailed image of a complex object can be mentally rotated as a whole (Cooper & Podgorny, 1976; Funt, 1983). Other results suggest limitations, such that rotation operates in a piecemeal manner for subsets of object structure (Folk & Luce, 1987; Just & Carpenter, 1985; for a review, see Khooshabeh, Hegarty, & Shipley, 2013).

Corresponding Author:

Yangqing Xu, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208
E-mail: xuy@u.northwestern.edu

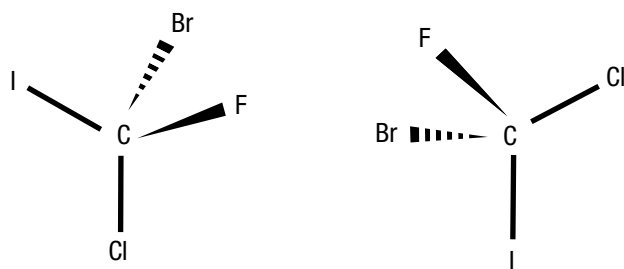


Fig. 1. Example of a mental rotation task in organic chemistry (adapted from Stieff, 2007). In this comparison task, students have to determine whether two molecules are identical (illustrated here) or have different structures. If the structures are different, they are mirror images, such that the molecules could not be spatially aligned even after imagined rotation.

But object shape is not the only type of visual information that must be transformed during mental rotation. For example, when chemistry students rotate a molecule to imagine it from a different perspective, they must keep visual feature information (in this case, symbols representing chemical elements, as in Fig. 1) attached to the correct parts of the object's rotated shape.

In the experiments reported here, we tested the capacity for keeping such features attached to the correct parts during mental rotation. All sample sizes were determined a priori to conform to conventional sample sizes in the visual cognition literature (e.g., Alvarez & Thompson, 2009; Wheeler & Treisman, 2002). We found a strikingly small capacity limit: Participants could keep only a single feature attached. This single feature appeared to remain attached by a single spotlight of attention that tracked the feature's position over time.

Experiment 1a

In Experiment 1a, we asked participants to rotate a simplified "molecule," a task requiring them to maintain attachments between features and moving parts. We found that they could keep only a single feature attached during mental rotation, and this capacity was much lower than capacity for the equivalent information in a static display.

Method

Participants. Twelve participants (18–35 years old) completed the experiment. Two of these participants were replacements for 2 others whose accuracy in the verbal suppression task (see Procedure) was less than 75%. All participants had normal or corrected-to-normal vision, were paid for their participation, and gave written consent.

Stimuli and apparatus. The experiment was controlled by a PC running SR Research Experiment Builder

(SR Research Ltd., Mississauga, Ontario, Canada). The display subtended $32.6^\circ \times 24.4^\circ$ at an approximate viewing distance of 56 cm and was presented on a 17-in. Dell E770S CRT monitor with a 75-Hz refresh rate and resolution of $1,024 \times 768$ pixels (33.6 pixels per degree). On each trial, participants were shown an abstracted four-bond molecule: a cross with four distinctly colored parts, each 10.7° long and 2.4° wide (see Fig. 2a). The four colors were randomly assigned to the four parts without replacement. The set of colors consisted of orange (RGB values: 233,122,0), green (RGB values: 0,166,0), aqua (RGB values: 0,136,233), and magenta (RGB values: 230,0,230), and the object was presented against a dark gray (RGB values: 80,80,80) background. In the initial display, any two adjacent parts formed a 90° angle, and the whole object was tilted either 10° clockwise (50% of trials) or 10° counterclockwise (50% of trials) from the cardinal orientation.

Procedure. In the object-rotation condition, trials were self-initiated and began with a cue animation showing a gray-scale pinwheel rotating for 2,400 ms; this rotation was accompanied by a continuous auditory clip of a mechanical sound mimicking a wheel rotating (Fig. 2a). The wheel rotated either clockwise or counterclockwise. Participants were instructed to think of the auditory clip as the sound that the wheel made while rotating and to remember the direction and rate of the wheel's rotation (a constant rate). They were informed that the cumulative amount of the wheel's angular rotation was irrelevant to the task. The to-be-rotated image was then presented statically for 500 ms, followed by a blank screen, which was presented for 800, 1,600, or 2,400 ms. Participants were instructed to pretend that a curtain dropped between them and the image, so that while the screen was blank, they would hear the mechanical sound but not see the rotating image. As soon as they heard the sound, they were to imagine the image rotating in the same direction and at the same rate as the wheel at the beginning of the trial, for as long as the mechanical sound played. They were told that at an unpredictable point, the "curtain would be raised" (though all display transitions were immediate), and another image would be revealed. Their task then was to indicate whether the image represented the same object with no feature swaps (i.e., with all colors attached to the correct postrotation parts) at the correctly rotated orientation.

In the control, no-rotation condition (see Fig. 2a), the cue was a static wheel with no sound. The four-part object was then presented for 500-ms, and participants were told to remember which colors were attached to which parts; no rotation was required. The delay intervals before the test image was presented were equivalent to those in the object-rotation condition.

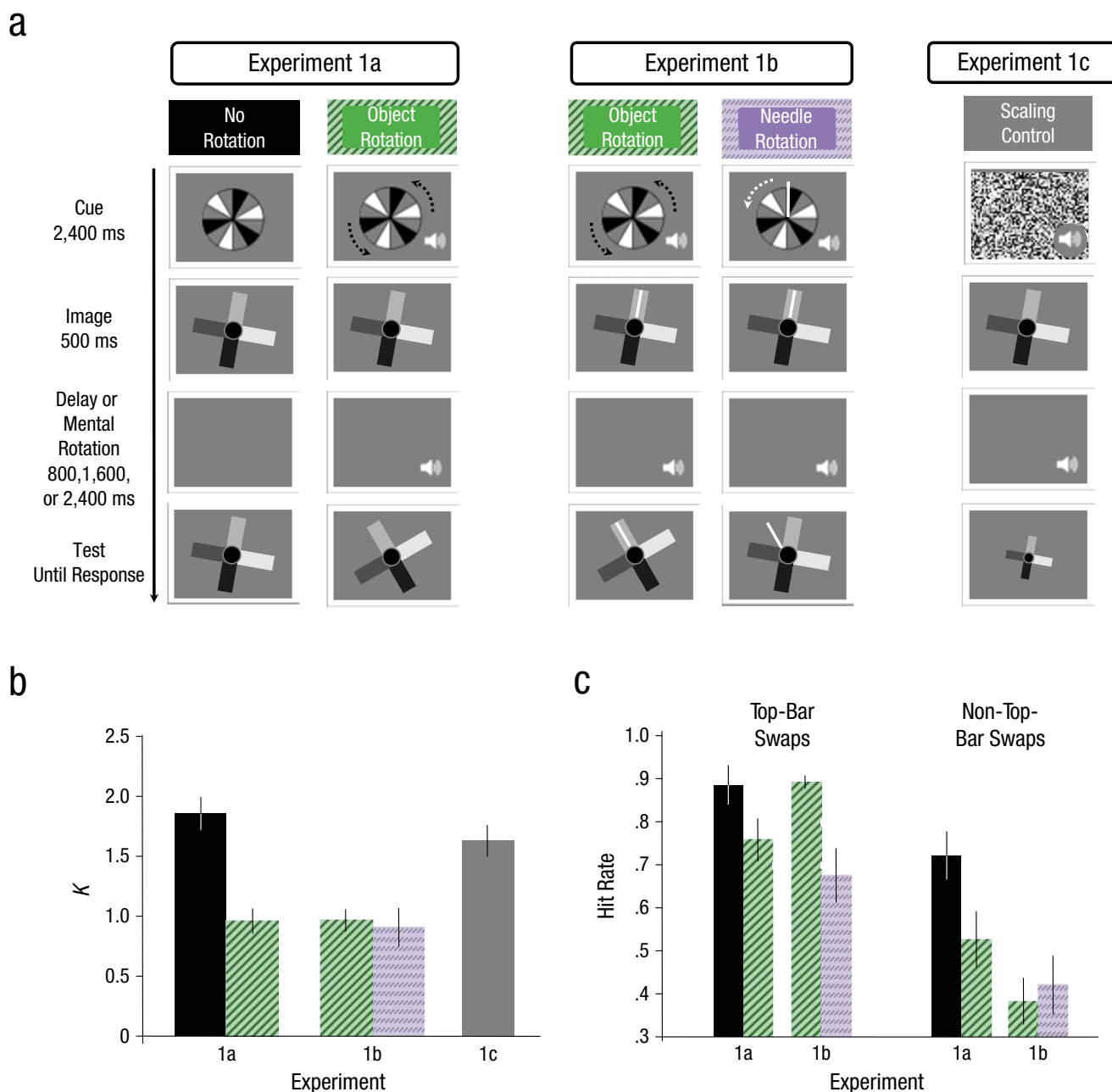


Fig. 2. Trial sequence and results of Experiments 1a, 1b, and 1c. In the object-rotation conditions of Experiments 1a and 1b (a), participants saw an object with four colored parts (shown here in gray). After the object disappeared, participants imagined the object rotating in the cued direction at the cued rate (a constant rate). They then indicated whether the image in the test display represented the correctly rotated object, with no feature swaps and the correct orientation. In the no-rotation condition, participants only detected potential feature swaps in the static object. In the needle-rotation condition, they mentally rotated a needle instead of the four-part object, and the test display contained either an incorrect needle orientation or a feature swap in the static object. In Experiment 1c (scaling control), participants imagined the object with four colored parts expanding or shrinking instead of rotating, and at test they indicated whether the image represented the correctly scaled object with no feature swaps. In all conditions, there was a verbal load, and the test image represented the initial object correctly on 50% of the trials. The graph in (b) presents the capacity estimates for all conditions in these experiments ($N = 12, 12$, and 11 for Experiments 1a, 1b, and 1c, respectively). The graph in (c) presents the hit rate in each condition, separately for when feature swaps did and did not involve the top part of the object. Error bars represent ± 1 SEM.

In both conditions, the test image represented the initial object correctly on 50% of the trials, and was

incorrect on the other 50% of the trials. Participants were given auditory feedback on their accuracy.

To isolate processing capacity for visual representations without the aid of verbal encoding, we included a verbal suppression task in both conditions. Prior to the cue in each trial, participants were presented with four nonrepeating consonants, and they were told to rehearse the letters mentally throughout the trial. Participants' memory for the letters was tested, unpredictably, at the end of 25% of the trials. Incorrect answers led to auditory feedback and a 3-s delay penalty, during which participants could not advance to the next trial.

Participants were told that they should monitor each test object for both inaccurate orientation and color swaps within the object. The experimenter guided each participant through several sample trials (equal numbers of all trial types within the design), providing verbal feedback on the participant's verbal responses. If the test image was an incorrect foil, the experimenter revealed whether it was an orientation foil or a feature-swap foil. After this interactive tutorial, the participant completed another set of self-paced practice trials before starting the actual experiment.

In the object-rotation condition, an incorrect foil was equally likely to be a feature-swap foil (i.e., correct orientation but two colored parts swapped) or an orientation foil (i.e., wrong orientation and no color swap). In the no-rotation condition, incorrect foils were always feature-swap foils. Task condition (clockwise object rotation, counterclockwise object rotation, or no rotation), test-image type (correct or incorrect), feature-swap foil (six possible foils), foil type (feature swap or orientation in the object-rotation condition; feature swap only in the no-rotation condition), and length of the rotation/memory interval (800, 1,600, or 2,400 ms) were fully crossed across 180 randomly ordered trials. Each participant was tested in five blocks of 36 trials. The entire experiment lasted approximately 60 min.

Results

The hit rate for orientation foils (correctly detecting that the test image was at a wrong orientation) was .58 in the object-rotation condition. The hit rate for feature-swap foils (correctly detecting a feature swap) was .64 in the object-rotation condition, but was .81 in the no-rotation condition. Using the hit rate for the feature-swap trials and the false alarm rate for the nonswap trials, we computed capacity (K) for the number of feature-part correspondences participants successfully stored. There were six possible swaps between pairs of colored parts. If only one out of four colored parts were remembered, three out of six possible swaps would involve a remembered part, and the resulting hit rate would be 3/6 plus the guessing rate (i.e., the false alarm rate,¹ the probability of reporting a swap when there was none). K can be

calculated from this relationship using a formula developed by Alvarez and Thompson (2009; see also Cowan, 2001; Pashler, 1988).

In the no-rotation condition, capacity was around 2 ($M = 1.86$, $SE = 0.15$, 95% confidence interval, or CI = [1.52, 2.19]). But capacity in the object-rotation condition was dramatically lower: Only a single feature could remain attached to a moving part ($M = 0.96$, $SE = 0.10$, 95% CI = [0.74, 1.18]), $F(1, 11) = 32.04$, $p < .001$, $\eta_p^2 = .74$ (Fig. 2b). An analysis of the results in terms of accuracy instead of capacity also revealed that performance was significantly lower in the object-rotation condition ($M = 76\%$, $SE = 2.1\%$, 95% CI = [71.3%, 80.7%]) than in the no-rotation condition ($M = 87.5\%$, $SE = 2.1\%$, 95% CI [82.9%, 92.1%]), $F(1, 11) = 19.63$, $p = .001$, $\eta_p^2 = .64$.

There was a marginal effect of the length of the blank interval on capacity, $F(2, 22) = 2.98$, $p = .07$. This effect was driven primarily by a marginally significant decrease in capacity on the 2,400-ms trials ($M = 1.21$, $SE = 0.15$, 95% CI = [0.88, 1.54]) compared with the 1,600-ms trials ($M = 1.64$, $SE = 0.13$, 95% CI = [1.34, 1.93]), $t(11) = 2.29$, $p = .04$. Average accuracy in the verbal suppression task was 90%.

Experiment 1b

Why does mental rotation destroy the ability to keep more than a single feature attached? Does the limit stem from the need to keep features attached to moving parts, or more simply from the need to keep features attached to parts, even if those parts are static? That is, if the no-rotation control condition of Experiment 1a were repeated, but participants were asked to rotate a *different* object, would the colors still become detached from the parts of the static object? In Experiment 1b, participants remembered the locations of four colors on a static cross, but simultaneously performed a rotation on a secondary object (a "needle" attached to the center of the cross). The results confirmed that feature capacity during mental rotation is deeply limited: Not only is it difficult to keep features attached as they move, but the attentional requirements of rotation even detach features from the memory representation of a static object.

Method

Participants. Twelve participants (18–35 years old) completed the experiment. All participants had normal or corrected-to-normal vision, were paid for their participation, and gave written consent.

Stimuli, apparatus, and procedure. The apparatus was identical to that used in Experiment 1a. The stimuli were the same except as noted here.

As in the object-rotation condition of Experiment 1a, the trial sequence in the needle-rotation condition consisted of presentation of a cue and then a four-part object (in this case with a gray needle attached to the center of the cross), followed by a blank interval and then a test image (Fig. 2a). The cue was a needle rotating either clockwise or counterclockwise on top of a static wheel; a continuous ticking sound accompanied the rotation. Participants were told that the four-part object would always remain static and the needle would rotate around its center independently. During the blank interval, the ticking sound was played again. Participants were asked to imagine the needle rotating in the same direction and at the same rate as the cue needle, for as long as the sound played. When the test display appeared, participants indicated whether the image represented the initial object (with all colors attached to the correct parts) and whether the needle was at the correct orientation.

This experiment also included a replication of the object-rotation condition of Experiment 1a, so that we could compare capacities. In this condition, the to-be-rotated object included a task-irrelevant needle, which always rotated with the four-part object as if it were glued to the top part of the object (Fig. 2a). When the test image appeared, participants indicated whether it represented the initial object (with all colors attached to the correct parts) at the correct orientation.

In both conditions, the test image was correct on 50% of the trials and was a foil on the other 50% of the trials. Participants were given auditory feedback indicating their accuracy. Participants also performed the same verbal suppression task as in Experiment 1a.

In the object-rotation condition, an incorrect foil was equally likely to be a feature-swap foil (i.e., the object was at the correct orientation, but two colored parts were swapped) or an orientation foil (i.e., the object was at the wrong orientation, but no colors were swapped). In the needle-rotation condition, an incorrect foil was equally likely to be a feature-swap foil (i.e., the needle was at the correct orientation, but two of the object's four colored parts were swapped) or an orientation foil (i.e., the needle was at the wrong orientation, but none of the object's colored parts were swapped). Condition (object-rotation or needle-rotation), rotation direction (clockwise or counterclockwise), test-image type (correct or incorrect), foil type (12 levels: 6 feature-swap foils, 6 orientation foils), and rotation period (800, 1,600, or 2,400 ms) were fully crossed across 288 randomly ordered trials. Each participant was tested in eight blocks of 36 trials. The entire experiment lasted approximately 90 min.

Results

The hit rate for orientation foils in the object-rotation condition (correctly detecting that the test image was at a

wrong orientation) was .65. The hit rate for orientation foils in the needle-rotation condition (correctly detecting that the needle was at a wrong orientation) was .78. The hit rate for feature-swap foils (detecting a feature swap when there was one) was .64 in the object-rotation condition and .55 in the needle-rotation condition.

Capacity for successfully storing feature-part correspondences was around 1 in both the needle-rotation condition ($M = 0.90$, $SE = 0.16$, 95% CI = [0.56, 1.25]) and the object-rotation condition ($M = 0.97$, $SE = 0.09$, 95% CI = [0.76, 1.17]; Fig. 2b); capacity did not differ between these conditions, $F(1, 11) = 0.54$, $p = .48$, n.s. Similarly, accuracy in the needle-rotation condition ($M = 77.2\%$, $SE = 2.4\%$, 95% CI = [72.0%, 82.5%]) was not significantly different from that in the object-rotation condition ($M = 77.0\%$, $SE = 1.6\%$, 95% CI = [73.4%, 80.6%]), $F(1, 11) = 0.02$, $p = .89$, n.s. There was no significant effect of the length of the rotation interval on either K or accuracy, $F_s(2, 22) < 0.88$, $p_s > .43$, n.s. Average accuracy in the verbal suppression task was 90.4%.

Experiment 1c

One might argue that the low capacity found in the mental rotation conditions (object rotation and needle rotation in Experiments 1a and 1b) was not due to the requirements of mental rotation; perhaps this impairment would occur for any difficult secondary task. To rule out this possibility, in Experiment 1c we tested a control condition in which participants performed a scaling task (imagined expansion or contraction) rather than a rotation task (imagined clockwise or counterclockwise rotation). Because scaling should not have the same requirements for attentional focusing and tracking that mental rotation does, and should allow participants to maintain attention on the entire object, we predicted that it would not disrupt feature-part binding. Indeed, performance on this scaling task showed little impairment compared with performance in the no-rotation condition of Experiment 1a, a result confirming that the low capacity observed in the mental rotation conditions of Experiments 1a and 1b did not stem from a general dual-task cost.

Method

Participants. Twelve participants (18–35 years old) completed the experiment. Data from 1 participant were excluded from all analyses because of accuracy less than 75% in the verbal suppression task. All participants had normal or corrected-to-normal vision, were paid for their participation, and gave written consent.

Stimuli, apparatus, and procedure. The stimuli, apparatus, design, and task were identical to those of the object-rotation condition in Experiment 1a except as

noted here. The cue at the beginning of each trial was an animation showing a white noise texture expanding or shrinking by 60% for 2,400 ms; this animation was accompanied by the same mechanical sound used in the object-rotation condition of Experiments 1a and 1b. Participants were told to encode the rate of the scaling. In this experiment, the blank interval following the initial display of the four-part object always had a duration of 2,400 ms. During this interval, participants were instructed to imagine the object expanding or contracting at the same rate as the texture surface, as if the image were glued to the surface. At test, they indicated whether the image represented the correct object, with no swaps of the colored parts, and in the correct scale (Fig. 2a).

Scaling direction (expand or contract), test-image type (correct or incorrect), and incorrect foil type (12 levels: 6 feature-swap foils, 6 scaling foils) were fully crossed across 144 randomly ordered trials. The entire experiment lasted approximately 60 min.

Results

The hit rate for scaling foils (detecting that the test image was incorrectly scaled) was .50, similar² to the hit rate for orientation foils in the object-rotation condition in Experiment 1a, $t(21) = 0.18$, n.s. However, the hit rate for feature-swap foils (detecting a feature swap when there was one) was far higher: .79.

Capacity (K) was significantly higher in this scaling control condition ($M = 1.63$, $SE = 0.12$, 95% CI = [1.35, 1.91]) than in the object-rotation condition in Experiment 1a ($M = 0.96$, $SE = 0.10$, 95% CI = [0.74, 1.18]), the object-rotation condition in Experiment 1b ($M = 0.97$, $SE = 0.09$, 95% CI = [0.76, 1.17]), and the needle-rotation condition in Experiment 1b ($M = 0.90$, $SE = 0.16$, 95% CI = [0.56, 1.25]), $ts(21) > 3.56$, $ps < .003$. Moreover, K in this scaling control condition did not differ from that observed in the no-rotation condition in Experiment 1a ($M = 1.86$, $SE = 0.15$, 95% CI = [1.52, 2.19]), $t(21) = 1.15$, $p = .26$, n.s. (see Fig. 2b). Average accuracy in the verbal suppression task was 92.3%.

Experiment 2

The consistent capacity limit of 1 suggests that participants preferentially selected one object part. But which part? We suspected that participants would preferentially attend to the object's top, given biases toward "top" framings over "bottom" framings in descriptions of spatial relationship (Clark & Chase, 1972; Tversky, 1975). The top of an object also appears to play a special role in thinking about clockwise and counterclockwise rotation: Turning a car's steering wheel clockwise makes the car go "right" because the wheel's top goes to the right (the

bottom goes to the left); similarly, the screwdriver rule of "lefty loosey, righty tighty" works only if one's focus is on the top of the screw. We predicted that participants would notice only those feature swaps involving the top part of the object.

Before collecting additional data, we looked back at performance in the object-rotation and needle-rotation conditions of Experiments 1a and 1b. Hit rates for swap detection were indeed higher when the top part was involved (around 75%, on average) than when it was not involved (near chance, on average; see the Results section later for statistical details). To verify this suggested focus on the object's top, in Experiment 2 we tracked the eye movements of a new set of participants as they completed a simplified mental rotation of two-part objects.

Method

Participants. Sixteen participants (18–35 years old) completed the experiment. All participants had normal or corrected-to-normal vision, were paid for their participation, and gave written consent.

Stimuli, apparatus, and procedure. Experiment 2 was similar to the object-rotation condition of Experiments 1a and 1b with the following differences. Eye movements were monitored by a table-mounted SR Research Eyelink 1000 Remote eye tracker (SR Research Ltd., Mississauga, Ontario, Canada). Also, the to-be-rotated object consisted of two colored parts (orange and green) instead of four (see Fig. 3a).

Results

The average accuracy for the task (averaged across feature-swap and orientation foils) was 78.8% across all participants. In the eye movement analyses, we excluded trials with no eye movements (8.2% of all trials), trials with no valid eye movements before the mental rotation period started (13% of trials), trials with initial saccades that did not start from the center of the screen (a virtual circle with a radius of approximately 2.6° around the fixation point; 0.7% of trials), and trials with final saccades that took place during a blink (3.1% of trials) or offscreen (0.04% of trials). The average percentage of the trials included in the saccade analyses was 74.9% across all 16 participants.

Given that the object in this experiment had only two parts (in contrast to the object in the previous experiments, which had four), observers should have been perfectly accurate at detecting swaps, because any swap had to involve both parts. To isolate K for feature swaps similarly to how we determined K in Experiments 1a and 1b, we used the correct-rejection rates in the trials with test

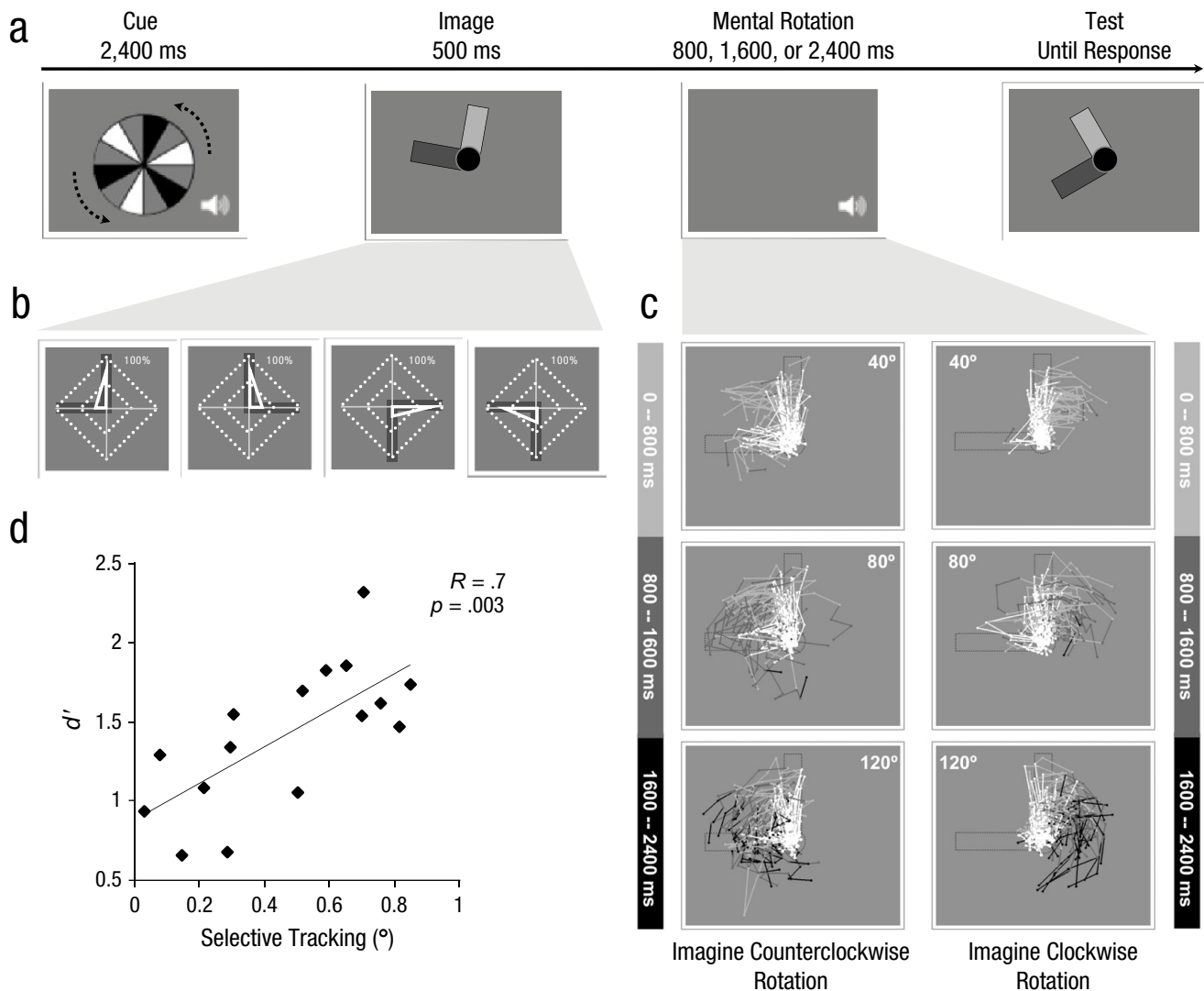


Fig. 3. Trial sequence and results of Experiment 2 ($N = 16$). As in Experiment 1, each trial began with a cue consisting of a rotating wheel accompanied by a continuous mechanical sound (a). The cue was followed by presentation of the to-be-rotated object. When this image disappeared, participants imagined it rotating in the cued direction and at the cued rate. Finally, they indicated whether the image in the test display represented the correctly rotated object, with no feature swaps and at the correct orientation. The test image was correct on 50% of the trials. The radar plots in (b) show the percentage of last saccades (i.e., immediately before the mental rotation period) in each quadrant for four object starting positions, indicated by the dark gray “arms.” In each plot, the outer dotted diamond indicates 100%, and the inner dotted diamond indicates 50%. The value represented along each of the four axes (top, bottom, left, and right) indicates the percentage of final saccades in the corresponding quadrant before mental rotation began. The illustrations in (c) show saccade traces for all participants in all of the trials in which participants responded correctly; traces are shown separately for trials in which the expected angle of mental rotation (based on the duration of the blank mental rotation interval) was 40° , 80° , and 120° . The shading of the traces indicates the time frames of the saccades (see the vertical bars on the side). White traces on the plots represent saccades that occurred after the trial started but before the mental rotation period started. The scatter plot (with best-fitting regression line) in (d) shows the relationship between accuracy in detecting orientation foils and degree of selective tracking (as measured by the average increment in eye movement angle for each degree of expected mental rotation).

images that correctly represented the rotated image (i.e., rates of correctly determining that the test image was in the correct orientation and had no feature swaps) and hit rates in the feature-swap trials. The average K (1.9 across 16 participants) was close to perfect, as predicted. The hit rate in the feature-swap trials was much higher in Experiment 2 (.93) than in the object-rotation conditions

of Experiments 1a and 1b (.64). Though this experiment was not designed to measure K , the results were consistent with our findings from Experiments 1a and 1b.

In order to test whether participants spontaneously selected one object part, we analyzed the position of the last fixation before the mental rotation period started for all the trials on which participants responded correctly

(80.3% of all trials with valid eye movements). We divided the screen into four virtual quadrants (top, bottom, left, and right) using two diagonal lines running from the corners and intersecting at the center of the screen. For each starting configuration of the object, we computed the average percentage of last fixations that fell in each quadrant.

Most participants spontaneously selected the topmost part of each object (Fig. 3b). In particular, when parts of the object were in the top and the left quadrants or the top and the right quadrants, the last saccades before mental rotation were more likely to land in the top quadrant ($M_s = 72.3\%$ and 73.5% , respectively) than in any of the other three quadrants, all $t_s(15) > 3.96$, $p_s < .002$. When parts of the object were in the bottom and the left quadrants or the bottom and the right quadrants, the last saccades before mental rotation were more likely to land in either the left quadrant ($M = 62.8\%$) or the right quadrant ($M = 74.6\%$) respectively, than in any of the other quadrants, all $t_s(15) > 2.61$, $p_s < .02$.

Most participants then kept their eyes locked to the imagined location of the initially selected part as it rotated, such that their eye movement paths followed the trajectory of mental rotation (Fig. 3c). We computed the centroid (i.e., the arithmetic mean position) of all saccades during the mental rotation period in each trial. In order to obtain the average position of each centroid with respect to the participant's initial selection, we collapsed the saccade locations across the trials by rotating the 2-D plane of the saccade positions in each trial such that the initial attentional selection always landed in the top quadrant. Because participants were cued to mentally rotate the object at a constant rate and the mental rotation intervals were 800, 1,600, and 2,400 ms, the expected extents of mental rotation were 40° , 80° , and 120° , respectively. We predicted that the positions of the centroids would land lower than and to the left of the selections when mental rotation was counterclockwise and lower than and to the right of the selections when mental rotation was clockwise. The results were consistent with this prediction. In the vertical direction, the centroids were more likely to be lower than the selections for both clockwise ($M = 71.8\%$) and counterclockwise ($M = 72.4\%$) mental rotation, $t_s(15) > 4.9$, $p_s < .001$. The direction of mental rotation made no difference, $t(15) = 1.24$, $p = .23$, n.s. In contrast, the direction of mental rotation did make a significant difference in the horizontal difference between the centroids and the selections, $t(15) = 10.16$, $p < .001$. For clockwise mental rotation, the centroids were more likely to be to the right of the selections than to the left ($M = 88.8\%$), $t(15) = 9.1$, $p < .001$, whereas for counterclockwise mental rotation, the centroids were more likely to be to the left of the selections than to the right ($M = 85.8\%$), $t(15) = 10.58$, $p < .001$ (Fig. 3c).

To evaluate whether eye movement paths were longer for trials with longer durations and greater rotation angles, we computed the eye movement angle (the angle between the first and last fixation during the mental rotation period with respect to the circle in the center of the object) in each trial (collapsing across the two rotation directions). These data were submitted to a repeated measures analysis of variance (degrees of freedom were Greenhouse-Geisser corrected for sphericity violations), with the expected angle of mental rotation (40° , 80° , or 120°) as the within-participants factor. There was a significant main effect of the expected angle of mental rotation, $F(1.3, 19.5) = 21.05$, $p < .001$, $\eta_p^2 = .58$. Paired comparisons revealed that the eye movement angle increased as the expected degree of mental rotation increased from 40° to 80° to 120° ($M_s = 27.5^\circ$, 45.7° , and 57.4° , respectively), all $t_s(15) > 3.5$, $p_s < .003$.

In addition, each participant's adherence to the part-tracking strategy was measured by computing the average increment in eye movement angle for each degree of expected mental rotation. Figure 3d reveals that this score was highly correlated to detection of the orientation foils ($r = .7$, $p = .003$; Fig. 3d).

We also analyzed the behavioral results of the object-rotation and needle-rotation conditions of Experiments 1a and 1b to verify that detection accuracy was far higher for swaps involving the top part of the object than for swaps not involving the top part (for which performance was near chance; Fig. 2c). In the object-rotation condition (collapsed across Experiments 1a and 1b), the hit rate for feature-swap foils was significantly higher when the swap involved the top part ($M = 82.4\%$, $SE = 2.6\%$, $95\% \text{ CI} = [76.9\%, 87.9\%]$) than when it did not involve the top part ($M = 45.6\%$, $SE = 4.3\%$, $95\% \text{ CI} = [36.7\%, 54.5\%]$), $F(1, 22) = 38.72$, $p < .001$, $\eta_p^2 = .64$. In the needle-rotation condition, the hit rate was again significantly higher when the swap involved the top part ($M = 67.6\%$, $SE = 6.1\%$, $95\% \text{ CI} = [54.1\%, 81.1\%]$) than when it did not involve the top part ($M = 42.1\%$, $SE = 7.2\%$, $95\% \text{ CI} = [26.3\%, 57.9\%]$), $F(1, 11) = 52.4$, $p < .001$, $\eta_p^2 = .83$.

General Discussion

Many types of representations contribute to mental transformation of an object's structure. Here, we focused on the ability to keep arbitrary visual features "glued" to particular parts of an object during mental rotation. The architecture of the human visual system appears badly suited for this task. Although participants could keep track of about two of these links when an object remained static, this capacity dropped to just one when that object rotated (Experiment 1a). This low capacity was apparent even when the object itself did not move, but when attention tracked a separate needle attached to the object

(Experiment 1b). These drops in capacity were not due to generalized dual-task costs. When participants were asked to transform an object by scaling, which should permit continued attention to an entire object (Experiment 1c), capacity was not impaired. Experiment 2 shows that the drops in capacity in Experiments 1a and 1b were likely due to the need to focus on and track individual parts with attention. In Experiment 2, participants tended to lock their gaze to one object part—the top—and track it. This was the same single part that the behavioral results indicate was encoded in Experiments 1a and 1b.

The ability to keep features glued to moving object *parts* appears to be subject to capacity limitations similar to those found in visual cognition research focusing on translation of separate *objects*. When asked to remember the feature correspondences for a set of simple colored objects (so as to detect swaps), participants can store two to three if the display remains static (see Brady, Konkle, & Alvarez, 2011), but this capacity drops substantially if the objects move (Horowitz et al., 2007; Saiki, 2003; Saiki & Miyatsuji, 2009). This impairment can even appear when participants are asked to track an unrelated set of objects while remembering a set of static objects (Fougnie & Marois, 2009)—a manipulation similar to our needle-rotation condition. This memory impairment is strongest when participants complete the demanding task of tracking moving objects with their attention; simply asking them to shift their focus of attention to a local region does not tend to strongly impair performance (e.g., Delvenne, Cleeremans, & Laloyaux, 2010; Gajewski & Brockmole, 2006; Johnson, Hollingworth, & Luck, 2008).

People may feel that they can holistically rotate a detailed representation, at least for a simple object. Our results suggest that this intuition is an illusion. Instead, mental rotation may rely on a deeply abstracted form of an object's shape, with visual feature information (e.g., colors) filled in on demand, pulled from other representations such as verbal coding (Stieff & Raje, 2010), whenever rotation is paused. This suggestion bridges proposals that view the processes that occur during mental imagery as being more pictorial (e.g., Kosslyn, Pinker, Smith, & Shwartz, 1979) and those that view these processes as being more abstract or propositional (e.g. Pylyshyn, 1981). The critical representation may be spatially depicted, but also abstracted to the point where mental rotation would be computationally feasible within known constraints of visual neurophysiology—perhaps transforming only the major axes depicting the core structure of an object (Marr & Nishihara, 1976).

Our behavioral and eye-tracking data suggest that the locus of attention may serve to mark “arrowheads” for the extracted major axes, allowing the computation of a new reference frame as the object rotates. Highlighting or adding information that provides an arrowhead for an

axis or a “front” for an object can improve mental rotation (Amorim, Isableu, & Jarraya, 2006; Hochberg & Gellman, 1977; Stull, Hegarty, & Mayer, 2009). This role for attention might also generalize to rotation in depth. Spatial selection can help pull the representation of a surface closer to the observer (Xu & Franconeri, 2012), and could play a similar role in 3-D rotation.

Although the present study tested the ability to keep features attached while mentally rotating an object through a noncanonical angle, there are also visual strategies for discrete rotation operations that avoid the type of continuous angular rotation required by our task. If we had asked for a 180° rotation of our stimuli, for example, observers could have simply swapped the positions of features along the vertical and horizontal axes. The situations in which such shortcuts can apply—and the capacity benefits that they may carry—are ripe for exploration.

Differences in such attentional strategies may also underlie many of the individual differences in mental rotation ability. Compared with low-spatial-ability participants, those with high spatial ability are more likely to solve mental rotation problems by focusing on specific parts of the objects, as evidenced by their gestures (Göksun, Goldin-Meadow, Newcombe, & Shipley, 2013). A beginning chemistry student who lacks domain-specific heuristics (Stieff, 2007) might attempt to rotate a molecular structure holistically—and fail to accurately transform the object. In contrast, professional chemists recognize the limitations of their visual system and rely on a set of analytical strategies, such as rotation of a single object part, combined with systematic verbal encoding of part labels (Stieff et al., 2014). Such strategic variations may also partially underlie individual differences that often correlate with gender (Peters, Lehmann, Takahira, Takeuchi, & Jordan, 2006; Stieff et al., 2014). Our results suggest that training students to efficiently deploy their limited visual resources, in combination with other visuo-spatial tools and external representations, could be a fruitful tool for science education.

Author Contributions

Both authors contributed to the study concept and study design. The experiments were programmed by Y. Xu. Testing and data collection were performed or supervised by Y. Xu. Y. Xu performed the data analysis under the supervision of S. L. Franconeri. Y. Xu drafted the manuscript, and S. L. Franconeri provided substantial revisions. Both authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Thidar Khine and Monika Lind for assistance with data acquisition and Mary Hegarty, Brandon Liverence, Andrew Lovett, Audrey Michal, Mike Stieff, Laura Stoughton, Satoru Suzuki, and David Uttal for helpful comments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was funded by National Science Foundation Grants BCS-1056730 and SBE-0541957 (to S. L. Franconeri) and by National Institutes of Health Grant T32 NS047987 (to Y. Xu).

Notes

1. The false alarm rate in the experiments we report in this article reflected the guess rate for both feature-swap and orientation judgments. A separate experiment (not reported here) showed that this rate was likely entirely driven by feature-swap decisions. In that experiment, features could be swapped, but the orientation of the test object was always correct, so all false alarms could be attributed to feature-swap judgments alone. Despite the absence of orientation foils, the false alarm rate (.17) was identical to the false alarm rates in Experiment 1a (.18) and Experiment 1b (.16).

2. Because this hit rate was (nonsignificantly) lower than the hit rates in our previous experiments, it seemed possible that the high capacity in this experiment was due to participants devoting a smaller share of processing resources to the scaling task than to the rotation task, perhaps because the scaling task was too difficult. We therefore conducted another experiment using a version of the scaling task with larger, easier-to-detect scaling differences. The hit rate for detecting scaling foils was high ($M = .62$) and statistically identical to the hit rates for detecting orientation foils in the object-rotation conditions of Experiments 1a ($M = .58$) and 1b ($M = .65$), both $ps > .61$. Yet the hit rate for detecting feature swaps remained high ($M = .74$), and the false alarm rate was low ($M = .1$). Capacity was also high; the value of K (1.42) was statistically identical to the value of K in the present experiment (1.63), $p = .32$, and was significantly higher than the values of K in the object-rotation condition of Experiment 1a (0.96) and the object-rotation condition of Experiment 1b (0.97), both $ps = .02$.

References

- Alvarez, G. A., & Thompson, T. W. (2009). Overwriting and rebinding: Why feature-switch detection tasks underestimate the binding capacity of visual working memory. *Visual Cognition*, 17, 141–159.
- Amorim, M. A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: "Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, 135, 327–347.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5), Article 4. Retrieved from <http://jov.arvojournals.org/article.aspx?articleid=2191865>
- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3, 472–517.
- Cooper, L. A., & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 503–514.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity [Target article and commentaries]. *Behavioral & Brain Sciences*, 24, 87–185.
- Delvenne, J. F., Cleeremans, A., & Laloyaux, C. (2010). Feature bindings are maintained in visual short-term memory without sustained focused attention. *Experimental Psychology*, 57(2), 108–116.
- Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 395–404.
- Fougnie, D., & Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Visual Cognition*, 17, 48–66.
- Funt, B. V. (1983). A parallel-process model of mental rotation. *Cognitive Science*, 7, 67–93.
- Gajewski, D. A., & Brockmole, J. R. (2006). Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin & Review*, 13, 581–587.
- Göksun, T., Goldin-Meadow, S., Newcombe, N., & Shipley, T. (2013). Individual differences in mental rotation: What does gesture tell us? *Cognitive Processing*, 14, 153–162.
- Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, 8, 280–285.
- Hegarty, M., Keehner, M., Cohen, C., Montello, D. R., & Lippa, Y. (2007). The role of spatial cognition in medicine: Applications for selecting and training professionals. In G. Allen (Ed.), *Applied spatial cognition* (pp. 285–315). Mahwah, NJ: Erlbaum.
- Hochberg, J., & Gellman, L. (1977). The effect of landmark features on mental rotation times. *Memory & Cognition*, 5, 23–26.
- Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, 69, 172–184.
- Johnson, J. S., Hollingworth, A., & Luck, S. J. (2008). The role of attention in the maintenance of feature bindings in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 41–55.
- Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, 92, 137–172.
- Khooshabeh, P., Hegarty, M., & Shipley, T. F. (2013). Individual differences in mental rotation: Piecemeal versus holistic processing. *Experimental Psychology*, 60, 164–171.
- Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). On the demystification of mental imagery. *Behavioral & Brain Sciences*, 2, 535–548.
- Marr, D., & Nishihara, H. K. (1976). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society B: Biological Sciences*, 200, 269–294.
- National Research Council. (2006). *Learning to think spatially*. Washington, DC: The National Academies Press.

- Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator*, 34(2), 29–35.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44, 369–378.
- Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four cross-cultural samples (N = 3367): Overall sex differences and the role of academic program in performance. *Cortex*, 42, 1005–1014.
- Pylyshyn, Z. W. (1979). The rate of “mental rotation” of images: A test of a holistic analogue hypothesis. *Memory & Cognition*, 7, 19–28.
- Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 88, 16–45.
- Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, 3(1), Article 2. Retrieved from <http://jov.arvojournals.org/Article.aspx?articleid=2158153>
- Saiki, J., & Miyatsuji, H. (2009). Estimated capacity of object files in visual short-term memory is not improved by retrieval cueing. *Journal of Vision*, 9(3), Article 16. Retrieved from <http://jov.arvojournals.org/article.aspx?articleid=2193412>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- Steiger, J. H., & Yuille, J. C. (1983). Long-term memory and mental rotation. *Canadian Journal of Psychology*, 37, 367–389.
- Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, 17, 219–234.
- Stieff, M., Dixon, B. L., Ryu, M., Kumi, B. C., & Hegarty, M. (2014). Strategy training eliminates sex differences in spatial problem solving in a STEM domain. *Journal of Educational Psychology*, 106, 390–402.
- Stieff, M., & Raje, S. (2010). Expert algorithmic and imagistic problem solving strategies in advanced chemistry. *Spatial Cognition and Computation*, 10, 53–81.
- Stull, A. T., Hegarty, M., & Mayer, R. E. (2009). Getting a handle on learning anatomy with interactive three-dimensional graphics. *Journal of Educational Psychology*, 101, 803–816.
- Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, 53, 433–441.
- Tversky, B. (1975). Pictorial encoding of sentences in sentence-picture comparison. *The Quarterly Journal of Experimental Psychology*, 27, 405–410.
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139, 352–402.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101, 817–835.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131, 48–64.
- Xu, Y., & Franconeri, S. L. (2012). The head of the table: The location of the spotlight of attention may determine the ‘front’ of ambiguous objects. *Journal of Neuroscience*, 32, 1408–1412.
- Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics*, 31, 201–209.
- Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, 20, 1–19.