Published in final edited form as:

Psychol Sci. 2015 August ; 26(8): 1241–1251. doi:10.1177/0956797615585002.

Capacity for visual features in mental rotation

Yangqing Xu and Steven L. Franconeri

Department of Psychology, Northwestern University

Abstract

Although mental rotation is a core component of scientific reasoning, we still know little about its underlying mechanism. For instance - how much visual information can we rotate at once? Participants rotated a simple multi-part 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 eyetracking 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 the 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

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 (Wai, Lubinski, & Benbow, 2009; Newcombe, 2010). Despite the importance of these abilities, students of both science and medicine have difficulty performing mental rotation (Hegarty, Keehner, Cohen, Montello, & Lippa, 2007; Stieff, Dixon, Ryu, Kumi, & Hegarty, 2013). Given the interest in training spatial thinking skills (National Research Council, 2006), researchers have sought strategies for improving them (e.g. Terlecki & Newcombe, 2005; see Uttal, Meadow, Tipton, Hand, Alden, Warren, & Newcombe, 2013, for review). But potential advances can be constrained by the fact that we do not yet fully understand the suite of underlying visuospatial tools that allow us 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, including not only visual information, but also motoric or other embodied representations (Zacks, 2008), verbal coding (Stieff & Raje, 2010), and long-term memory (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 3D blocks (Shepard & Metzler, 1971; Yuille & Steiger, 1982), geometric 2D drawings (Pylyshyn, 1979), and 2D random polygons (Cooper & Podgorny, 1976; Folk & Luce, 1987). Some of these results suggest a virtually unlimited capacity for shape transformation, where a detailed image of a complex object can be mentally rotated as a whole (Cooper & Podgorny, 1976; Funt, 1983). Other results suggest limitations, with rotation operating in a ‘piecemeal’ manner for subsets of object structure (Just & Carpenter, 1985; Folk & Luce, 1987; see Khooshabeh, Hegarty, & Shipley, 2013, for review).

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

Here we test the capacity for keeping such features attached to the correct parts during mental rotation. We found a strikingly small limit: participants could keep only a single feature attached. This single feature appears to remain attached by a single spotlight of attention that tracks the feature’s position over time.

Experiment 1A

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

Method

Participants—Twelve participants (18–35 yr) completed the experiment. Two of these participants were replacements for two others with accuracy of less than 75% in the verbal suppression task. All sample sizes in the present study were determined a priori conforming to conventional sample sizes in the visual cognition literature (e.g. Alvarez & Thompson, 2009; Wheeler & Treisman, 2002). All participants had normal or corrected-to-normal vision, were paid for their participation, and gave written consent.

Stimuli and Apparatus—The experiments were controlled by a PC running SR Research Experiment Builder. The display subtended $32.6^{\circ} \times 24.4^{\circ}$ at an approximate viewing distance of 56cm, on a 17" Dell E770S CRT monitor with a 75-Hz refresh rate, and 1024×768 pixel resolution (33.6 pixels per degree). Participants mentally rotated an abstracted four-bond molecule (a cross with four distinctly colored parts, Figure 1B). On each trial the four colors were randomly assigned to the four parts without replacement. The color parts were drawn in orange (233,122,0), green (0,166,0), aqua (0,136,233), and magenta (230,0,230), against a dark grey (80,80,80) background. The length and width of the parts was 10.7° and 2.4° respectively. In the initial display, any two adjacent parts

formed a 90-degree angle and the whole object was tilted either 10° clockwise or 10° counter-clockwise (50% chance) from the cardinal orientation.

Procedure—In the mental rotation condition, self-initiated trials began with a cue animation showing a grayscale pinwheel rotating for 2400 ms accompanied by a continuous auditory clip displaying a mechanical sound mimicking a wheel rotating (Figure 1B). The wheel rotated either clockwise or counter-clockwise. Participants were instructed to think of the auditory clip as the sound that the wheel makes while rotating and to remember the direction and rate of the wheel's rotation. The cumulative amount of the wheel's angular rotation was irrelevant to the task. The to-be-rotated image then appeared for 500 ms followed by a blank screen, during which the participants were instructed to pretend that a curtain had dropped between them and the image, so that they would hear the mechanical sound but the rotating image was invisible. As soon as they heard the sound, they imagined the image rotating in the same direction and rate as the wheel at the beginning of the trial, for as long as the mechanical sound was playing. 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, at which point participants indicated whether the image represented the same object with no feature swaps (with all colors attached to the correct post-rotation parts) at the correctly rotated orientation. In a control condition, the cue was a static wheel with no sound. Participants remembered which colors were attached to which parts during equivalent delay intervals, but no rotation was required. In both conditions, there was a 50% chance of same/different trials and they were given an auditory feedback indicating accuracy.

To isolate the processing capacity for visual representations without the aid of verbal encoding, both conditions also included a verbal suppression task. Prior to the cue animation of each trial, participants were presented with four non-repeating consonants. They were told to rehearse the letters mentally throughout the trial and were tested on their memory of these letters at the end of the trial unpredictably on 25% of the trials. Incorrect answers led to auditory feedback and a three second delay penalty.

Participants were told that they should monitor both for inaccurate orientation changes as well as the color-swaps within the object. The experimenter guided each participant through several sample trials with an equal number of each trial type within the design, with verbal feedback for the participant's verbal responses to the sample trials. 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, participants then completed another set of self-paced practice trials before they start the actual experiment.

In the mental rotation condition, the incorrect foil image was equally likely to be a feature-swap foil (the object at the correct orientation with a swap of two colored parts) or a wrong-orientation foil (the object at a wrong orientation with no swap). In the no-rotation condition, the incorrect foil was always a feature-swap foil. The task condition (clockwise, counter-clockwise mental rotation, or no-rotation), the test image type (correct or incorrect), the feature-swap foil (6 possible foils), the foil type (swap or rate in the mental rotation condition; swap only in the no-rotation condition), and the length of the rotation/memory

interval (800, 1600, 2400 ms) were fully crossed across 180 randomly ordered trials. Each participant was tested in 5 blocks of 36 trials. The entire experiment lasted approximately 60 minutes.

Results—The hit rate for rate-foil trials (correctly detecting that the test image was at a wrong orientation) was 0.58 in the mental rotation condition. The hit rate for feature-swap foil trials (detecting a feature swap when there was one) was 0.64 in this mental rotation condition, but was 0.81 in the no-rotation condition. Based on the hit rate for the feature-swap trials, and the false alarm rate for the non-swap trials, we computed capacity (K) for the number of feature-part correspondences participants successfully stored. There were 6 possible swaps between pairs of the color parts. If only 1 out of 4 color parts were remembered, 3 out of 6 possible swaps would involve that remembered part, resulting in a hit rate of 3/6 plus the guessing rate (i.e. false alarm rate¹, the probability of reporting a swap for the non-swap trials). This relationship can be expressed with K as a dependent variable using a formula developed by Alvarez & Thompson (2009; see also, Cowan, 2001; Pashler, 1988).

In the no-rotation condition, when no rotation was required, capacity was around 2 (M=1.86, SE=0.15, 95% CI [1.52, 2.19]). But capacity in the mental rotation condition was drastically lower – only a single feature could remain attached to a moving part (M=0.96, SE=0.10, 95% CI [0.74, 1.18]), $F(1, 11)=32.04$, $p < 0.001$, $\eta^2=0.74$ (Figure 1C). An analysis of the results in terms of accuracy instead of capacity also revealed that performance in the mental rotation condition (M=76%, SE=2.1, 95% CI [71.3%, 80.7%]) was significantly lower than that in the no-rotation condition (M=87.5%, SE=2.1%, 95% CI [82.9%, 92.1%]), $F(1, 11)=19.63$, $p=0.001$, $\eta^2=0.64$.

There was a marginal effect of the length of the blank interval on capacity, $F(2, 22)=2.98$, $p=0.07$, primarily driven by a marginally significant decrease in capacity from the 1600 ms trials (M=1.64, SE=0.13, 95% CI [1.34, 1.93]) to the 2400 ms trials (M=1.21, SE=0.15, 95% CI [0.88, 1.54]), $t(11)=2.29$, $p=0.04$. 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

¹False alarm (FA) rates in the present study reflected the guess rate for both the feature-swap judgment and the orientation judgment, a separate experiment (not reported here) shows that this FA rate is likely entirely driven by the feature-swap decision. Features could swap but rotation angle was always correct, so false alarm rates were isolated to the feature-swap judgment. The FA rate for feature-swap alone (0.17) was identical to FA rates in E1A (0.18) and E1B (0.16).

²Because this hit rate is (non-significantly) lower than the values in our previous experiments, it could invite the possibility that the high capacity in this experiment was due to participants devoting a smaller share of processing resources to the scaling task, perhaps because it is too difficult. We therefore replicated these results in a version of the scaling task with larger scaling differences that are easier to detect. The hit rate for the scaling judgment was high (M=0.62), identical to the previous experiments (1A and 1B). Yet, hit rate for detecting feature swaps remained high (M=0.74), false alarm rates low (M=0.1), and capacity high (K=1.42), statistically identical ($p=0.32$) to the present experiment (K=1.63), and significantly higher than K values in the two rotation conditions, (K=0.96; K=0.97), both p 's = 0.02.

to rotate a *different* object, would the colors still become detached from the static object? In Experiment 1B, participants remembered the locations of four colors on a static cross, but performed an equivalent rotation on a secondary object (a ‘needle’ attached to the center of the cross). The results confirm that feature rotation capacity 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.

Participants

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

Stimuli and Apparatus

The to-be-rotated object was identical to that in Experiment 1A, with the addition of a grey ‘needle’ attached to the center of the cross (Figure 1A).

Procedure

In the needle-rotation condition, the cue was a needle rotating either clockwise or counterclockwise on top of a static wheel, accompanied by a continuous ticking sound. 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 rate as the cue needle, for as long as the sound was playing. When the test display appeared, participants indicated whether the static image represented the same 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 mental rotation condition in order to compare their capacities. In this condition, a task-irrelevant needle was still present, but would always rotate with the four-part object as if it were glued to the top part of the initial object. When the test image appeared, participants indicated whether the image represented the same object (with all colors attached to the correct parts) at the correct orientation. In both conditions, there was again a 50% chance of same/different trials and participants were given auditory feedback indicating the accuracy. The verbal suppression task was identical to that in Experiment 1A.

In the object-rotation condition, the incorrect foil image was equally likely to be a feature-swap foil (the object at the correct orientation with a swap of two of four colored parts) or an orientation-foil (the object at a wrong orientation with no feature swap). In the needle-rotation condition, the incorrect foil was equally likely to be a feature-swap foil (the needle at the correct orientation with a swap of two of four colored parts) or an orientation-foil (the needle at a wrong orientation with no feature-swap between parts). The condition (object-rotation or needle-rotation), rotation direction (clockwise or counterclockwise), test image type (correct or incorrect), incorrect foil type (12 levels) and rotation period length (800, 1600, 2400 ms) were fully crossed across 288 randomly ordered trials. Each participant was tested in 8 blocks of 36 trials. The entire experiment lasted approximately 90 minutes.

Results

The hit rate for the object rotation judgment (correctly detecting that the test image was at a wrong orientation) was 0.65 in the object mental rotation condition. The hit rate for the needle rotation judgment (correctly detecting that the needle was at a wrong orientation) was 0.78 in the needle rotation condition. Hit rate for the feature-swap judgment (detecting a feature swap when there was one) was 0.64 in the mental rotation condition and 0.55 in needle rotation condition.

Capacity for the number of feature-part correspondences participants successfully stored was around 1 in both the needle rotation ($M=0.90$, $SE=0.16$, 95% CI [0.56, 1.25]) and the replication of the object mental rotation condition ($M=0.97$, $SE=0.09$, 95% CI [0.76, 1.17]), $F(1, 11)=0.54$, $p=0.48$, ns. Similarly, accuracy in the needle rotation condition ($M=77.2\%$, $SE=2.4\%$, 95% CI [72%, 82.5%]) was not significantly different from that in the mental rotation condition ($M=77\%$, $SE=1.6\%$, 95% CI [73.4%, 80.6%]), $F(1, 11)=0.02$, $p=0.89$, ns. There was no significant effect of the length of the rotation interval, $F(2, 22)=0.88$, $p=0.43$, ns. Average accuracy in the verbal suppression task was 90.4%.

Experiment 1C

One might argue that the low capacity in the mental rotation was not due to the requirements of mental rotation – perhaps this impairment would occur for any difficult secondary task. To rule out this possibility, Experiment 1C tested a control condition that replaced the rotation task (an imagined clockwise or counter-clockwise rotation) with a scaling task (an imagined expansion or contraction). Because scaling should not have the same requirements for attentional focusing and tracking that are found in the rotation and needle rotation conditions, and should allow participants to maintain attention on the entire object, we predicted that it would not disrupt feature-part binding. Indeed, this scaling task led to little impairment compared to the no-rotation condition in E1A, confirming that the low capacity obtained in the mental rotation condition in E1A and 1B did not stem from a general dual task cost.

Participants

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

Procedure

The task was identical to the mental rotation condition in Experiment 1A with the following differences. The cue at the beginning of the trial was an animation showing a white noise texture expanding or shrinking by 60% for 2400 ms, accompanied by the same mechanical sound used in Experiment 1A and Experiment 1B. The participants were told to encode the rate of the scaling. During the 2400ms blank interval, the participants were instructed to imagine the four-part object expanding or contracting in the same rate as the texture surface, as if the image were glued to the surface. At the test display, they indicated whether the

image represented the correct object, with no feature swaps of the color parts, and in the correct scale (Figure 1b).

The design was otherwise identical to previous experiments. The scaling direction (expand or contract), the test image type (correct or incorrect), and the incorrect foil type (12 levels: 6 swaps foils, 6 scaling foils) were fully crossed across 144 randomly ordered trials. The entire experiment lasted approximately 60 minutes.

Results

The hit rate for scaling judgment (detecting that the test image was incorrectly scaled) was 0.50, similar² to that for the rotation judgment in the mental rotation condition in Experiment 1A, $t(21)=1.8$, ns. But the hit rate for feature-swap judgment (detecting a feature swap when there was one) was far higher: 0.79.

Capacity (K) in this scaling control ($M=1.63$, $SE=0.12$, 95% CI [1.35, 1.9]) was significantly higher than in the mental rotation conditions in Experiment 1A ($M=0.96$, $SE=0.10$, 95% CI [0.74, 1.18]) and 1B ($M=0.97$, $SE=0.09$, 95% CI [0.76, 1.17]), and in the needle rotation condition in Experiment 1B ($M=0.90$, $SE=0.16$, 95% CI [0.56, 1.25]), $t's(21)>3.56$, $p's<0.003$. Moreover, K in the current scaling control 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]), $t(21)=1.15$, $p=0.26$, ns. (See Figure 1C) 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, based on biases toward 'top' framings over 'bottom' framings in spatial relationship descriptions (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 clockwise makes it 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' only works if one's focus is on the top of the screw. We predicted that participants would only notice feature swaps involving the top part of the object.

Before collecting additional data, we first looked back to performance in the mental rotation conditions of Experiments 1A & 1B. Swap detection hit rates were indeed higher when the top part was involved (around 83% on average), relative to when the top part was not involved (near chance, on average; see the results section below for statistical details). To verify that these behavioral results suggest a focus on the object's top, in Experiment 2 a new set of participants were eye-tracked as they completed a simplified mental rotation of simpler objects with two parts (Figure 2A).

Participants

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

Stimuli and Apparatus

Experiment 2 was similar to the mental rotation condition Experiment 1A and 1B with the following differences. Eye movements were monitored by a table-mounted SR-Research Eyelink 1000 Remote eyetracker. The to-be-rotated object consisted of two colored parts (orange and green) instead of four (Figure 2A).

Results

The average accuracy for the task (averaged across hit rate for color swap judgment and rotation judgment) was 78.8% across all participants. In the following 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%), 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) (0.7%), and trials with final saccades within a blink (3.1%) or offscreen (0.04%). The average percentage of the trials included in the saccade analyses was 74.9% across all 16 participants.

Because there were only two parts in the object in this experiment (in contrast to the previous experiment which had four), observers should be perfect at detecting swaps, because any swap must involve both parts. To isolate feature-swap K similarly to Experiment 1A and 1B, we used the CR in the ‘correct trials’ and hit rate in the ‘feature swap trials’. The average was close to perfect ($K=1.9$ across 16 participants) as predicted. The Hit rate in the ‘feature swap trials’ in Experiment 2 (0.93) was much higher than that in Experiment 1 with 4 colored parts (0.64). Though this experiment was not designed to measure K , the results were consistent with our findings from Experiment 1.

In order to test whether people spontaneously selected one object part, we analyzed the position of the last fixation before the mental rotation period started for all the correct mental rotation trials (80.3% of all valid trials). We divided the screen into four virtual quadrants (top, bottom, left, right) by two diagonal lines that run from the corners and intersect at the center of the screen. For each starting configuration of the object, we computed the average percentage of the last fixation right before the mental rotation period that fell in each quadrant.

Most participants spontaneously selected the topmost part of each object (Figure 2B). 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 most likely to land in the top quadrant ($M=72.3\%$, $M=73.5\%$, respectively) than any other three quadrants, all $t's(15)>3.96$, $p's<0.002$. 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 most likely to land in either the left quadrant ($M=62.8\%$) or the right quadrant ($M=74.6\%$) respectively, than any other three quadrants, all $t's(15)>2.61$, $p's<0.02$.

Most participants then kept their eyes locked to the imagined location of this part as it rotated, with eye movement paths following with the trajectory of mental rotation (Figure 2C). 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 positions of the centroids

with respect to the participant's initial selections in the corresponding trial, we collapsed across the trials by rotating the 2d-plane of the saccade positions in each trial such that the initial selection always landed in the top quadrant. Following our hypothesis, we predicted that the positions of the centroids should land on the lower left side of the selections for counter-clockwise rotations, and on the lower right side of the selections for clockwise rotations. 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 counter-clockwise ($M=72.4\%$) mental rotation, $t(15) > 4.9$, $p's < 0.001$. The direction of mental rotation made no difference, $t(15)=1.24$, $p=0.23$, ns. In contrast, the direction of mental rotation did make a significant difference in the horizontal difference between the centroid and the selection, $t(15)=10.16$, $p < 0.001$. For clockwise mental rotation, the centroids were more likely to be to the right of the selection ($M=88.8\%$), $t(15)=9.1$, $p < 0.001$, whereas for counter-clockwise mental rotation, the centroids were more likely to be to the left of the selection ($M=85.8\%$), $t(15)=10.58$, $p < 0.001$ (Figure 2C).

To evaluate whether eye movement paths were longer for trials with longer duration and rotation angle, we computed the eye movement angle (the angle between the first and last fixation with respect to the object center circle) in each trial (collapsing across both rotation directions). The saccade angle data were submitted to a repeated-measures ANOVA (degrees of freedom were Greenhouse-Geisser corrected for sphericity violations), with the expected angle of mental rotation (40° , 80° , 120°) as the within-participant factor. There was a significant main effect of the expected angle of mental rotation, $F(1.3, 19.5)=21.05$, $p < 0.001$, $\eta^2=0.58$. The saccade angle increased as the expected degree of mental rotation increased from 40° , 80° , to 120° ($M=27.5^\circ$, 45.7° , 57.4° , respectively), all $t's(15) > 3.5$, $p's < 0.003$.

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 intended mental rotation. Figure 2D reveals that this score was highly correlated to detection of the angular offset foils ($r=0.7$, $p=0.003$) (Figure 2D).

We also returned to the behavioral results of the mental rotation and needle rotation conditions of Experiments 1A and 1B to verify that swap detection accuracy was far higher for swaps involving the top part of the object, compared to near-chance performance when it did not (Figure 1D). In Experiment 1A, swap detection hit rate was significantly higher when the color-swap involved the top part ($M = 82.4\%$, $SE = 3.6\%$, $95\% \text{ CI } [74.6\%, 90.2\%]$) compared to when the color-swap did not involve the top part ($M = 62.5\%$, $SE = 5.5\%$, $95\% \text{ CI } [50.4\%, 74.6\%]$), $F(1, 11)= 7.02$, $p = 0.02$, $\eta^2 = 0.39$. In Experiment 1B, hit rates were again significantly higher when the color-swap involved the top part ($M = 78.2\%$, $SE = 3.3\%$, $95\% \text{ CI } [71\%, 85.5\%]$) compared to when the color-swap did not involve the top part ($M = 40.3\%$, $SE = 6.3\%$, $95\% \text{ CI } [26.5\%, 54.1\%]$), $F(1, 11)= 93.86$, $p < 0.001$, $\eta^2 = 0.90$.

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. While participants could keep around 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 even apparent when the object itself did not move, but when attention tracked a separate 'needle' attached to the object (Experiment 1B). These drops in capacity are 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), capacities returned to high levels. Experiment 2 shows that drops in capacity in 1A and 1B are likely due to the need to focus and track individual parts with attention. Participants tended to lock their gaze to one object part (systematically the top) and track it. This was the same single part that was encoded in 1A and 1B as shown by the behavioral results.

The ability to keep features glued to moving object *parts* appears to be subject to similar capacity limitations as those found in visual cognition literatures focusing on translation of separate *objects*. When asked to remember the feature correspondences for a set of simple colored objects (being ready to detect 'swaps'), participants can store 2–3 if the display remains static (see Brady, Konkle, & Alvarez, 2011). But this capacity drops substantially as soon as the objects move (Horowitz, Klieger, Fencsik, Yang, Alvarez, & Wolfe, 2007; Saiki, 2003; Saiki, 2009). This impairment can even appear when participants are asked to track an *unrelated* set of objects while storing a set of static objects (Fougnie & Marois, 2009), in a manipulation similar to our 'needle' condition. These memory impairments are 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. Gajewski & Brockmole, 2006; Johnson, Hollingworth & Luck, 2008; Delvenne et al., 2010).

People may feel that they can holistically rotate a detailed representation, at least for simple objects. 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 (Steiff & Raje, 2010), whenever rotation is paused. This type of abstract structural representation bridges proposals of more pictorial presentation (e.g. Kosslyn, Pinker, Smith, & Shwartz, 1979) and more abstract or propositional representations (e.g. Pylyshyn, 1981) during mental imagery. 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 eyetracking data suggest that the locus of attention may serve to mark the 'arrowhead' of this major axis, allowing the computation of a new reference frame as the object rotates. Highlighting or adding information that provides an axis 'arrowhead' or object 'front' 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 3D rotation.

While the present study tested the ability to keep features attached while mentally rotating an object for a non-canonical angle, there are also visual strategies for discrete rotation operations that avoid the type of continuous angular rotation required by our task. If we asked for a 180-degree rotation of our stimuli, an observer could just swap the positions of features along the vertical and horizontal axes. The situations where 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 to low-spatial ability participants, high-spatial ability participants are more likely to solve mental rotation problems by focusing on specific parts of an object, 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., 2013). Such strategic variations may also partially underlie individual differences that often correlate with gender (Peters, Lehmann, Takahira, Takeuchi, & Jordan, 2006; Stieff, et al., 2013). Our results suggest that training students to efficiently deploy their limited visual resources, in combination with other visuospatial tools and external representations, could be a fruitful tool for science education.

References

- Alvarez GA, Thompson TW. Overwriting and rebinding: Why feature-switch detection tasks underestimate the binding capacity of visual working memory. *Visual Cognition*. 2009; 17(1–2): 141–159.
- Amorim MA, Isableu B, Jarraya M. Embodied spatial transformations: “body analogy” for the mental rotation of objects. *Journal of Experimental Psychology: General*. 2006; 135(3):327–347. [PubMed: 16846268]
- Brady TF, Konkle T, Alvarez GA. A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*. 2011; 11(5):4, 1–34. [PubMed: 21617025]
- Clark HH, Chase WG. On the process of comparing sentences against pictures. *Cognitive psychology*. 1972; 3(3):472–517.
- Cooper LA, Podgorny P. Mental transformations and visual comparison processes: effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance*. 1976; 2(4):503–514. [PubMed: 1010999]
- Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*. 2001; 24(1):87–185. [PubMed: 11515286]
- Delvenne JF, Cleeremans A, Laloyaux C. Feature bindings are maintained in visual short-term memory without sustained focused attention. *Experimental Psychology*. 2010; 57(2):108–116. [PubMed: 20178921]
- Folk MD, Luce RD. Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance*. 1987; 13(3):395–404. [PubMed: 2958588]

- Fougnie D, Marois R. Attentive tracking disrupts feature binding in visual working memory. *Visual cognition*. 2009; 17(1–2):48–66. [PubMed: 19609460]
- Funt BV. A parallel-process model of mental rotation. *Cognitive Science*. 1983; 7(1):67–93.
- Gajewski DA, Brockmole JR. Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin & Review*. 2006; 13(4):581–587. [PubMed: 17201355]
- Göksun T, Goldin-Meadow S, Newcombe N, Shipley T. Individual differences in mental rotation: what does gesture tell us? *Cognitive processing*. 2013; 14(2):153–162. [PubMed: 23423638]
- Hegarty M. Mechanical reasoning by mental simulation. *Trends in cognitive sciences*. 2004; 8(6):280–285. [PubMed: 15165554]
- Hegarty, M.; Keehner, M.; Cohen, C.; Montello, DR.; Lippa, Y. The role of spatial cognition in medicine: Applications for selecting and training professionals. In: Allen, G., editor. *Applied spatial cognition*. Mahwah, NJ: Lawrence Erlbaum Associates; 2007. p. 285–315.
- Hochberg J, Gellman L. The effect of landmark features on mental rotation times. *Memory & Cognition*. 1977; 5(1):23–26. [PubMed: 21331862]
- Horowitz TS, Klieger SB, Fencsik DE, Yang KK, Alvarez GA, Wolfe JM. Tracking unique objects. *Perception & psychophysics*. 2007; 69(2):172–184. [PubMed: 17557588]
- Johnson JS, Hollingworth A, Luck SJ. The role of attention in the maintenance of feature bindings in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*. 2008; 34(1):41–55. [PubMed: 18248139]
- Just MA, Carpenter PA. Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological review*. 1985; 92(2):137–172. [PubMed: 3887449]
- Khooshabeh P, Hegarty M, Shipley TF. Individual differences in mental rotation: Piecemeal versus holistic processing. *Experimental psychology*. 2013; 60(3):164–171. [PubMed: 23128584]
- Kosslyn SM, Pinker S, Smith GE, Shwartz SP. On the demystification of mental imagery. *Behavioral and Brain Sciences*. 1979; 2(04):535–548.
- Marr D, Nishihara HK. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B Biological Sciences*. 1976; 200(1140):269–294.
- Newcombe NS. Picture this: increasing math and science learning by improving spatial thinking. *American Educator*. 2010; 34(2):29–35.
- Pashler H. Familiarity and visual change detection. *Perception & Psychophysics*. 1988; 44(4):369–378. [PubMed: 3226885]
- Peters M, Lehmann W, Takahira S, Takeuchi Y, Jordan K. Mental rotation test performance in four cross-cultural samples (n=3367): overall sex differences and the role of academic program in performance. *Cortex*. 2006; 42(7):1005–1014. [PubMed: 17172180]
- Pylyshyn ZW. The rate of “mental rotation” of images: A test of a holistic analogue hypothesis. *Memory & Cognition*. 1979; 7(1):19–28. [PubMed: 431403]
- Pylyshyn ZW. The imagery debate: Analogue media versus tacit knowledge. *Psychological review*. 1981; 88(1):16–45.
- Saiki J. Feature binding in object-file representations of multiple moving items. *Journal of Vision*. 2003; 3(1):6–21. [PubMed: 12678621]
- Saiki J, Miyatsuji H. Estimated capacity of object files in visual short-term memory is not improved by retrieval cueing. *Journal of vision*. 2009; 9(3):16, 1–15. [PubMed: 19757955]
- Shepard RN, Metzler J. Mental rotation of three-dimensional objects. *Science*. 1971; 171:701–703. [PubMed: 5540314]
- Steiger JH, Yuille JC. Long-term memory and mental rotation. *Canadian Journal of Psychology*. 1983; 37:367–389. [PubMed: 6640445]
- Stieff M. Mental rotation and diagrammatic reasoning in science. *Learning and instruction*. 2007; 17(2):219–234.
- Stieff M, Dixon BL, Ryu M, Kumi BC, Hegarty M. Strategy training eliminates sex differences in spatial problem solving in a stem domain. *Journal of Educational Psychology*. 2013; 106(2):390–402.

- Stieff M, Raje S. Expert algorithmic and imagistic problem solving strategies in advanced chemistry. *Spatial Cognition & Computation*. 2010; 10(1):53–81.
- Stull AT, Hegarty M, Mayer RE. Getting a handle on learning anatomy with interactive three-dimensional graphics. *Journal of educational psychology*. 2009; 101(4):803–816.
- Terlecki MS, Newcombe NS. How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*. 2005; 53(5–6):433–441.
- Tversky B. Pictorial encoding of sentences in sentence–picture comparison. *The Quarterly Journal of Experimental Psychology*. 1975; 27(3):405–410.
- Uttal DH, Meadow NG, Tipton E, Hand LL, Alden AR, Warren C, Newcombe NS. The malleability of spatial skills: a meta-analysis of training studies. *Psychological bulletin*. 2013; 139(2):352–402. [PubMed: 22663761]
- Wai J, Lubinski D, Benbow CP. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*. 2009; 101(4):817–835.
- Wheeler ME, Treisman AM. Binding in short-term visual memory. *Journal of Experimental Psychology: General*. 2002; 131(1):48–64. [PubMed: 11900102]
- Xu Y, Franconeri SL. The head of the table: The location of the spotlight of attention may determine the ‘front’ of ambiguous objects. *Journal of Neuroscience*. 2012; 32(4):1408–1412. [PubMed: 22279225]
- Yuille JC, Steiger JH. Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics*. 1982; 31(3):201–209. [PubMed: 7088661]
- Zacks JM. Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of cognitive neuroscience*. 2008; 20(1):1–19. [PubMed: 17919082]

Fig 1.

A. Example organic chemistry comparison task (adapted from Stieff, 2007).

B. Experiment 1 trial sequence. In the mental rotation condition, participants saw an object with four colored parts (shown here in grey). In Exp 1A and 1B, after the object disappeared, they imagined the object rotating in the cued direction and rate. At the test display, they indicated whether the image represented the correctly rotated object – 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, participants rotated a needle instead of the four-part object, and the test display could contain either an incorrect needle orientation or a feature swap to the static object. In Exp 1C, they imagined the object expanding or shrinking instead of rotating, and at test they indicated whether the image represented the correctly scaled object with no feature swaps. All conditions had a verbal load, and 50% same or different trials.

C. Capacity estimates for the no rotation, mental rotation, needle rotation, and mental scaling task. On average, participants were only able to maintain one feature-part binding

during both mental rotation and needle rotation. Capacity for scaling was similar to capacity for no rotation. **D.** In the mental rotation and needle rotation task, participants performed at chance when the feature-swap did not involve the top part, suggesting that participants were only tracking the top part in the rotation transformation. Error bars represent ± 1 SEM. (N=12, 12, 11)

Figure 2.

Experiment 2 shows that participants tend to select, and track, a single object part - the top, **A**. Trial sequence. Similar to Experiment 1, when the image disappeared, participants imagined the two-part image rotating in the cued direction and rate. At the test display, they indicated whether the image represented the correctly rotated object – no feature swaps and the correct orientation. There was a 50% chance that the test image was the correct image, **B**. Percentage of last saccades in each quadrant immediately before the mental rotation period for different image starting positions. Participants selectively attended to one of the two parts in the to-be-rotated image and they typically selected the topmost part. **C**. Eyetracking revealed that as the rotation operation progressed, people systematically guided their gaze along the imaged track of the selected part. **D**. Performance in detecting angular offset foils was highly correlated with the degree of selective tracking. (N=16)