Cognitive Psychology 167 (2026) 101852



Contents lists available at ScienceDirect

Cognitive Psychology

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



Conflict between representational formats and procedures constrains visual tracking

Qihan Wu, Jonathan I. Flombaum *

Johns Hopkins University, United States

ARTICLE INFO

Keywords:
Perception
Representation
Tracking
Composite
Format
Vision

ABSTRACT

Computer-assisted design software, such as PowerPoint, groups separate elements to create structured objects whose parts can be manipulated relative to the whole. Might visual perception perform a similar operation, grouping scene elements and representing them with internal reference to the emergent structure? The proposal is central yet perennially controversial in cognitive neuroscience and philosophy of mind. In a series of seven preregistered tracking experiments, we identify internally referenced representations in human vision by relying on the logic that, as in computer-assisted design, a representation will make certain procedures easy to access while limiting others. Participants tracked a target among identical nontargets organized into geometric shapes. When the shapes rotated, performance was high, suggesting the use of internally referenced, shape-centered representations that naturally support rotational transformations. In contrast, performance declined in translation conditions, where constituent items moved independently, revealing a mismatch between the internally structured representation and the requisite, externally oriented tracking procedure. This effect replicated across diverse stimuli, including occluded, rounded, and irregular shapes, and was reduced when structure-inducing cues were disrupted. These findings identify representational format and its compatibility with task procedures as an important determinant of visual performance.

1. Introduction

In his seminal book, David Marr (1982) argued that a full understanding of vision must characterize the nature of the representations it uses and the processes that operate on them. Simon and Newell (1971) shared a broader sentiment, arguing that a theory of intelligence must account for how representations are constructed and used. Not everyone agrees. James Gibson (2014) famously advocated for *direct* perception, in which environmental affordances are perceived without a need for mental representations. Rodney Brooks (1990, p. 5) succinctly agreed with a quip: "The world is its own best model."

The disagreement persists. A group of philosophers argued recently that "Mental representations remain the central posits of psychology after many decades of scrutiny. However, there is no consensus about the representational format(s) of biological cognition" (Quilty-Dunn et al., 2023). This owes, in part, to the difficulty of seeing directly the components of representations—symbols, variables, models—in the brain, if it is possible at all (Quilty-Dunn et al., 2023). Representations and their parts are functional units, concrete in the same sense as representations in computer graphics.

A revealing property of representations as functional units (in the mind, in computers, and anywhere else) is that they can conflict

* Corresponding author.
E-mail address: flombaum@jhu.edu (J.I. Flombaum).

https://doi.org/10.1016/j.cogpsych.2026.101852
Received 22 December 2025; Received in revised form 2 October 2026; Accepted 1 November 2026
0010-0285/© 2026 Published by Elsevier Inc.

with procedures, a point that Marr (1982, p. 21) made as follows: "The issue is important, because how information is represented can greatly affect how easy it is to do different things with it." When a representational format and a procedure conflict, the task will be difficult: error prone, slow, or both, compared to a representation more appropriately suited to the task and its attendant procedures. Here we exploit a representational conflict to demonstrate the presence of shape-centered—i.e., internally referenced—representations in visual perception. We use the term *reference frame* here in a relatively narrow geometric sense. We are not referring to the broader distinctions among retinotopic, egocentric, allocentric, or other spatial reference frames studied extensively in vision science and spatial cognition. Instead, our concern is with the internal geometry of a represented object or configuration: whether the locations of its parts are described relative to axes, centers, or relations intrinsic to that structure.

An internally referenced representation of a shape (Vannuscorps et al., 2022) or an object (McCloskey et al., 2006) situates parts relative to coordinates that remain static—a Platonic or canonical ideal, so to speak—so that locating the parts in the present traffics through variables that describe how the whole is arranged relative to an external reference frame. Representational conflicts should arise when a procedure demands external coordinates, but the native representation supplied is internally referenced.

What do you see in Fig. 1A for example, sixteen differently colored discs organized as a square, or a square comprising sixteen differently colored discs? The distinction may seem like one without a difference. Indeed, representations are descriptions, and language affords flexibility and productive potential for redescription. But in a domain-specific context, like predicting the outcome of a rotation for instance, these representations will perform differently because of their underlying formats. "Sixteen colorful discs organized as a square" would include the location and color of each disc with reference to some external reference-frame, as shown in Fig. 1B. That the arrangement is consistent with a square would be a further inference given the recorded locations (arguably, not even a part of the representation if one adheres to a strong version of direct perception). If one believes that the arrangement is incidental, that the discs might move independently and that some or all of them would need to be tracked, then representing them independently and with external reference is a good idea. If instead one wants to know what the square will look like when rotated 90° clockwise, this same representation is poorly suited.

In contrast, a representation of "a square comprising sixteen differently colored discs," as shown in Fig. 1C, is well suited to predicting the outcome of a rotation. A rotation procedure can be applied to the parts of the representation that describe how the square is situated in the present relative to an external reference frame, so that the anticipated stimulus is the same square from before, now with its top side parallel to the external vertical instead of the external horizontal. Transforming an entire shape is not intrinsically easier or more straightforward than changing one of its parts. But once the discs are represented as parts of a square, their positions within the square remain stable under rotation. It would therefore be difficult to use such a representation to simulate the appearance of the square if each disc moved three positions counterclockwise, for example. Such translation around the perimeter imposes a different demand because the discs change their positions within the represented structure itself. The procedure therefore requires updating the very part-to-whole relations that the internally referenced representation otherwise treats as stable. Anyone who has worked with PowerPoint™ or similar slide-making software will appreciate how grouping the discs in Fig. 1A produces a new representation in the software, making certain procedures available, rotation among them. Once grouped, rotating such a stimulus is a procedure that inherently refers to a past version of itself.

Note that we have supplied arbitrary representational details, including Cartesian coordinates, labels such as "Top" and "Right," and cardinally oriented axes. We are not committed to any of these specifications as they are isomorphic with alternatives. What matters for current purposes is the presence or absence of square-internal reference. The representation in Fig. 1C is internally referenced because it describes the positions of the colored discs relative to an underlying square structure, an object that really does

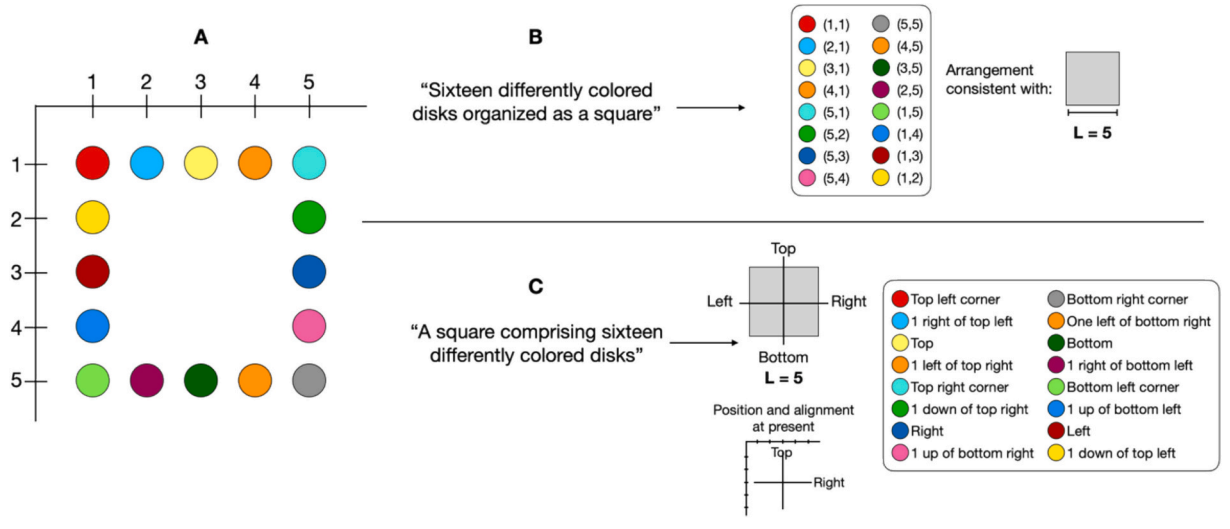


Fig. 1. (A) A stimulus. (B) A representation of the stimulus as "sixteen differently colored discs organized as a square." (C) A representation of the stimulus as a "square comprising sixteen differently colored discs."

not exist on its own but only as the product of the arrangement of the individuals. Yet it is intuitive to describe the individuals in terms that refer to this collective arrangement, and it is intuitive to localize them by first describing how the whole arrangement is positioned and oriented relative to an external frame in the present.

The logic of the current experiments follows accordingly. If a stimulus naturally elicits an internally referenced representation, then it should accommodate performance in some tasks and hinder performance in others. We therefore employ a concrete tool used in contemporary cognitive science to infer the structure of internal representations. Recent work has used this logic to argue for specific representational formats in domains including magnitude representation and visual role–filler structure (Hafri et al., 2024; Odic et al., 2024).

Related evidence comes from our recent work on the motion-silencing illusion (Wu & Flombaum, 2024). There, local visual changes became harder to detect when the changing elements were perceived as parts of a rotating object, including cases in which the rotational organization was supplied perceptually rather than locally. We interpreted this as evidence that rotation can induce an object-centered representation in which local elements are encoded relative to the larger structure. The present experiments build on the same idea, asking whether such a representational format facilitates some procedures while interfering with others.

In the experiments that follow, we arranged red discs to define basic shapes. Although the discs were not physically connected, we expected them to elicit shape-internal representations, that is, that vision would naturally describe the positions of the discs in terms of their locations on the overall structure, rather than with reference to external coordinates. This kind of representation should perform well if a participant is asked to track one disc while the whole stimulus rotates, since tracking in this case can rely on a procedure that updates the orientation of the whole shape relative to the external frame, without needing to update the internally referenced description of the target position. In contrast, we anticipated that performance should be comparatively poor when someone is asked to

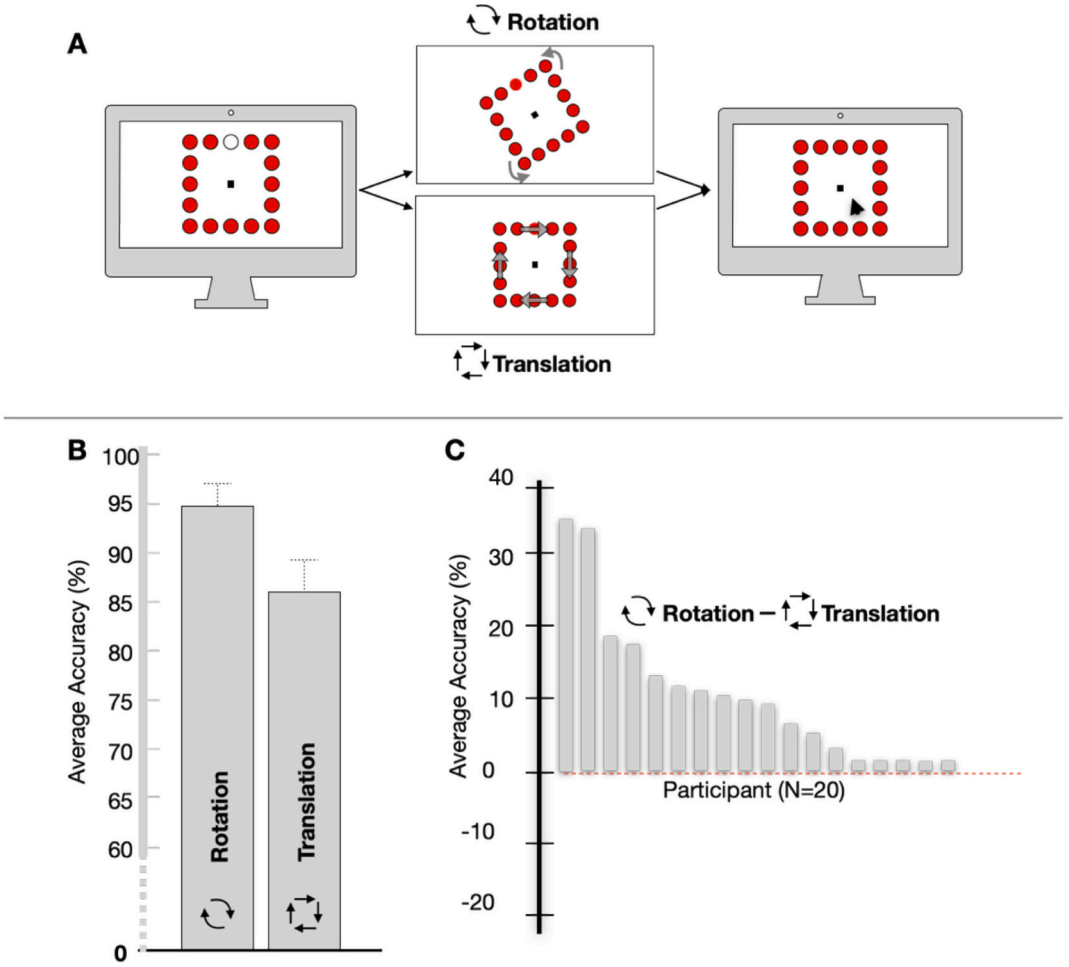


Fig. 2. (A) Stimuli and task (not to scale) for Experiment 1. Trials always begin with sixteen red discs organized to form a square. One randomly and temporarily turns white to identify it as a target. Participants track the target through a motion period, identifying it by clicking when motion ceases. The motion is Rotation in half of trials and Translation in the remaining half. (B) Average tracking performance for the two motion conditions. (C) Difference in average tracking performance (Rotation – Translation) for each of the twenty participants. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

track an individual while the constituent discs translate, occupying successive positions on the perimeter of the shape. To test the prediction, we employed an exceedingly simple tracking task. At the start of each trial, one of the red discs turned white temporarily, to identify it as a target, and the participant tracked that target during a motion period, reidentifying it at the end.

2. Experiment 1: Square rotation versus square translation

The stimulus in Experiment 1 was a set of sixteen red discs organized around a square (Fig. 2A). Although the discs were not physically connected, we expected a square-centered native representation. By our hypothesis, this kind of representation should perform well if a participant is asked to track one disc while the whole square rotates. At the start of each trial, one of the sixteen red discs turned white temporarily, to identify it as a target, and the participant tracked that target during a motion period, reidentifying it at the end. The experiment can be viewed here.

2.1. Methods

All the experiments described were preregistered with aspredicted.org. The registration for Experiment 1 can be found here. A repository of code and data has been included with this submission, to be made public following review.

2.1.1. Participants

All participants were recruited through an online system for students to receive course-related credit. Demographic information was not collected since it was not related to the hypothesis. Participants completed the experiment on a personal device. The instructions and sign-up page asked that they use a laptop or desktop computer, not a mobile device. The protocol was approved by the Johns Hopkins University Homewood Campus IRB.

In each of the six experiments we aimed to collect data from 20 participants. The sample size was determined based on pilot experiments conducted prior to preregistration. Any participant with a correct response rate below 10% overall was excluded and replaced. Such data were discarded automatically and never viewed. The same criterion was applied to all experiments.

We tested a total of 22 undergraduates for Experiment 1, with 2 excluded as described.

2.1.2. Stimuli and procedures

All Parameters are reported in pixels because the experiments were run online. In Experiment 1, each trial included 16 discs arranged as a square in the center of the screen. See Fig. 2. Each disc had a radius of 17.5 pixels and the square had a side length of 380 pixels. At the start of a trial, one of the discs was white, indicating it was the target, while the rest were red. The target was chosen randomly. At the start of each trial, all the discs remained stationary for 2000 ms. Then the white disc turned red, and the discs remained stationary for another 500 ms. Next, the discs moved for 5000 ms. When motion ceased, an arrow cursor appeared in the display. Participants were instructed to use the mouse to click on the disc that they believed to be the target.

Trials included two motion types, randomly and equally distributed over the course of a testing session. In the Rotation condition, the virtual square formed by the discs rotated around its center in either a clockwise or counterclockwise direction. The rotation speed was 120°/s, so that each disc completed one revolution in 3000 ms. Consequently, linear speed for a target in the Rotation condition varied depending on the disc's position on the square. An item occupying one of the four corners traced a circular path with a circumference of 1688 pixels, corresponding to a linear speed of 0.56 pixels/ms. An item occupying the midpoint of one of the square's four edges traced a circular path with a circumference of 1193 pixels, corresponding to a linear speed of 0.40 pixels/ms. Items occupying the six intermediate positions traveled at speeds between these extremes. In the Translation condition, each disc translated in one cardinal direction, turning corners as needed to stay on the perimeter of the virtual square. Translation speed was set so that each disc completed one lap in 3000 ms, corresponding to a linear speed of 0.50 pixels/ms. Each participant completed 120 trials, with Translation and Rotation trials distributed equally and randomly throughout the session.

Results were computed after excluding any individual trial with a reaction time less than 200 ms or greater than 8000 ms. 200 ms is a typical cutoff time in such experiments. 8000-ms trials were excluded based on pilot experiments prior to preregistration, which suggested that some subjects might step away from the task for long durations, a challenge with online experiments. We analyzed results from 2346 trials in Experiment 1.

2.2. Results

As we predicted, when the square rotated during the motion period, tracking performance was very high on average (>94%; Fig. 2B). In contrast, tracking performance was significantly lower, on average, when the target and nontarget discs underwent Translation rather than Rotation around the perimeter of the virtual square; planned one-tailed *t*-test, $t(19) = 4.45$, $p = 1.37*10^{-4}$, $d = 1.00$. Fig. 2C plots the average difference between the two conditions for each of 20 participants.

2.3. Discussion

The key result is the challenge to tracking under Translation. We predicted the effect on the reasoning that tracking operations for such motion cannot rely on internally referenced representations, since the translating discs do not remain in stable positions in a square-centered frame. Therefore, the target in a Translation trial must be tracked independently, via externally referenced

coordinates.

The difficulty is surprising, nonetheless, because this is known to be an easy task when it includes a single target that moves independently among independently moving nontargets. It has been studied extensively in the context of the Multiple Object Tracking paradigm (MOT; Pylyshyn & Storm, 1988), and it has been modeled successfully using externally referenced coordinates as input (e.g., Ma & Huang, 2009; Upadhyayula & Flombaum, 2020). But here we predicted that it would be more difficult because the square arrangement of discs would elicit a square-centered representation that would conflict with the tracking procedure needed to track independent motion (though it would support tracking through rotation).

One alternative is that square rotation benefits from categorical or verbal coding rather than from perceptual tracking per se. For example, a participant might encode the target as occupying a particular position relative to a corner or side of the square and use that description to guess or recover the target after a lapse. Experiments 5 and 6 will return directly to this possibility, first by eliminating the categorical landmarks supplied by the square and then by equating the final guessing probability across motion conditions.

3. Experiment 2: Rounded corners

In Experiments 2 and 3, we controlled for an alternative: that the cause of worse tracking performance under Translation has something to do with difficulty tracking a target as it turns a corner. Experiment 2 therefore included stimuli with rounded corners, so that the shape made by the discs was square-like (Fig. 3A). The experiment can be viewed here.

3.1. Methods

The registration for Experiment 2 can be found here.

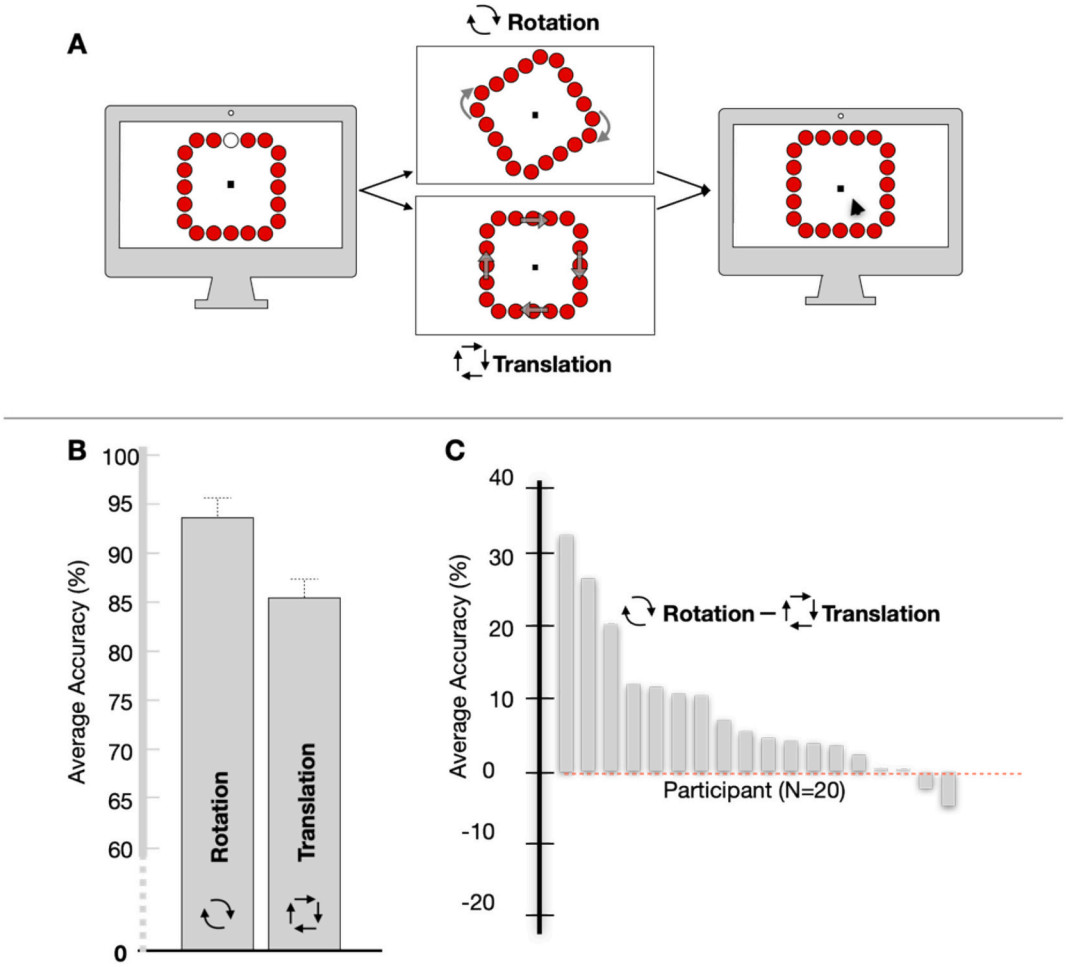


Fig. 3. (A) Stimuli and task (not to scale) for Experiment 2. The experiment is identical to Experiment 1 except that the squares include round corners (and 20 instead of 16 discs). (B) Average tracking performance for the two motion conditions. (C) Difference in average tracking performance (Rotation – Translation) for each of the twenty participants.

3.1.1. Participants

20 participants were tested for Experiment 2 (no exclusions).

3.1.2. Stimuli and procedures

Experiment 2 was identical to Experiment 1 except as follows. Each trial included 20 discs. The discs were organized to form a square, with 5 per side. The sides were arranged so that the corners were absent; in other words, the corner disc was removed from an equivalent square made from 7 discs per side. Each disc had a radius of 17.5 pixels. The square had a side length of 460 pixels, with the 5 discs occupying 340 pixels. We analyzed results from 2384 trials in Experiment 2.

3.2. Results

The results were very similar to Experiment 1 (Figs. 3B and C): Performance was significantly better in Rotation trials compared to Translation, $t(19) = 3.80$, $p = 6.06*10^{-4}$, d = 0.85. Fig. 3C plots the average difference between the two conditions for each of the 20 participants.

3.3. Discussion

Experiment 2 suggests that translation including rectilinear corners is an unlikely cause for the comparative tracking deficit for Translation compared to Rotation in Experiment 1. Experiment 3 adduces further evidence with a different stimulus type.

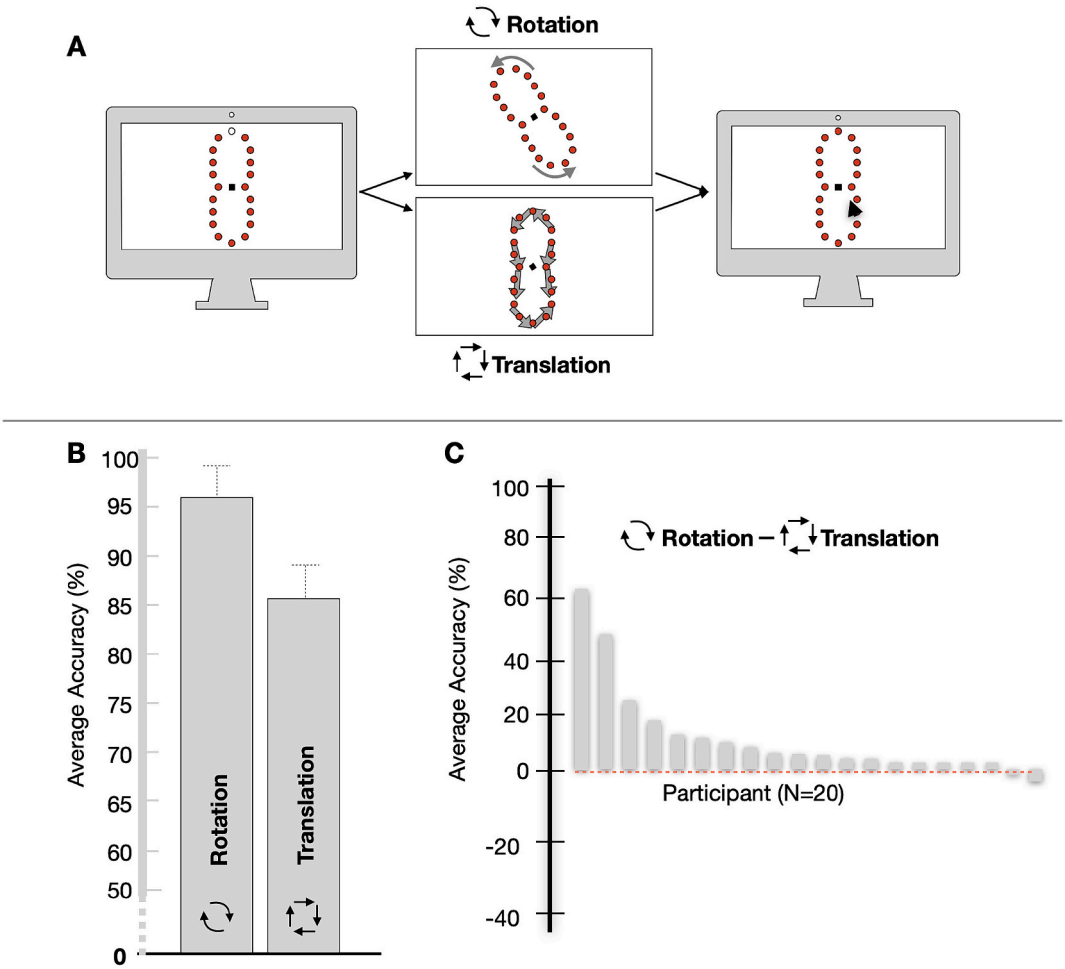


Fig. 4. (A) Stimuli and task (not to scale) for Experiment 3. The experiment is identical to Experiment 1 except that the squares were replaced with incomplete figure eights. (B) Average tracking performance for the two motion conditions. (C) Difference in average tracking performance (Rotation – Translation) for each of the 20 participants.

4. Experiment 3: A non-rectilinear shape

Experiment 3 moved away from a square to a shape lacking in straight sides altogether. The experiment required a shape that could rotate and that also would provide for a Translation condition, where constituent elements could move along its perimeter. As we describe in the context of Experiment 5, a circle does not meet these provisos. Neither does an ellipse, because translation around the perimeter of an ellipse induces a percept of circular rotation in a plane that is tilted relative to the image plane. The interpretation of the stimulus is that the ellipse is a perspectival distortion consistent with a circle in three dimensions (Rokers et al., 2006). We therefore generated an object that can be roughly described as an incomplete figure eight (Fig. 4A; a demonstration of the experiment can be viewed here).

4.1. Methods

The registration for Experiment 3 can be found here.

4.1.1. Participants

20 participants were tested for Experiment 3 (no exclusions).

4.1.2. Stimuli and procedures

Experiment 3 was identical to Experiment 1, except as described. Instead of a square, the discs were arranged on the perimeter of an

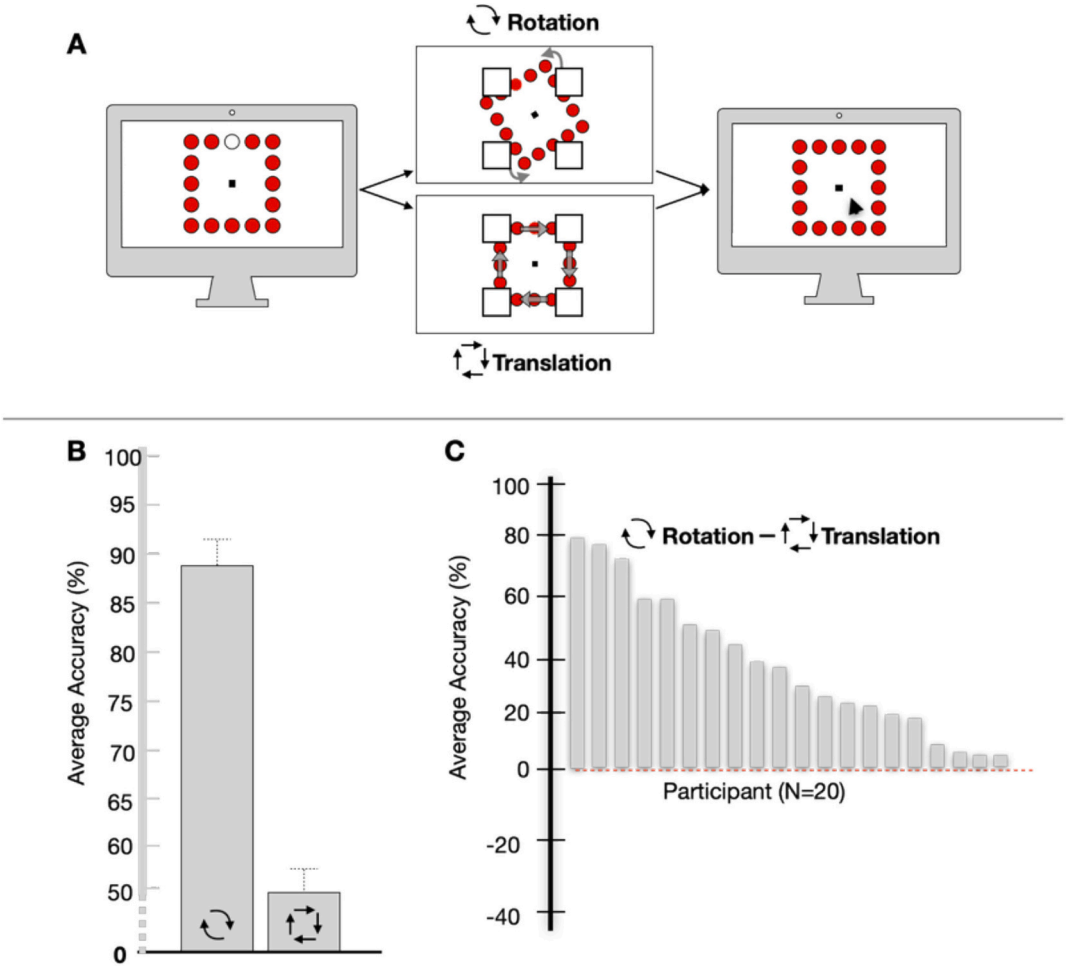


Fig. 5. (A) Stimuli and task (not to scale) for Experiment 4. The experiment is identical to Experiment 1 except for the presence of occluders shown in the middle panels. In the experiment, the occluders slide smoothly over the stimulus before a motion period begins, and smoothly out of the way at the end of each motion period. In the experiment, the occluders were also present during the stimulus presentation and test, but positioned adjacent to the stimuli. (B) Average tracking performance for the two motion conditions. (C) Difference in average tracking performance (Rotation – Translation) for each of the 20 participants.

incomplete figure eight. The discs either translated or rotated as in Experiment 1. In both motion conditions, each disc took 4.03 s to complete one lap. We analyzed results from 2393 trials in Experiment 3.

4.2. Results

In the Rotation condition, again, tracking performance was significantly better, on average, than in the Translation condition, $t(19) = 2.90$, $p = 4.58*10^{-3}$, d = 0.65; Fig. 4B. Fig. 4C shows the performance differences on a per-participant basis.

4.3. Discussion

Experiment 3 thus demonstrates that a tracking advantage for the rotation of a perimeter compared to translation around a perimeter is not a by-product of straight edges and turning at corners.

5. Experiment 4: Tracking through occlusion

An important feature of visual tracking ability under common circumstances is robustness in the face of amodal periods of occlusion. Babies and children understand that objects persist when they move behind occluding surfaces (Spelke et al., 1995). In the Multiple Object Tracking paradigm, complete occlusion does not impair performance (Scholl & Pylyshyn, 1999), and observers direct attention to moments of occlusion (Flombaum et al., 2008). In these experimental contexts, tracking takes place by utilizing externally referenced coordinates, and the tracked objects move independently from one another and from any additional moving objects in the scene. We therefore sought to determine whether tracking a single target that is part of a rotating ensemble is also a process that is robust to occlusion. Is tracking through occlusion a procedure that must rely on externally referenced coordinates, or can it utilize internally referenced ones, perhaps advantageously when the overall arrangement of elements comports with a shape-centered representation? Experiment 4 therefore included occluders in the display during the motion period (Fig. 5A; a demonstration can be viewed here).

5.1. Methods

The registration for Experiment 4 can be found here.

5.1.1. Participants

22 participants were tested for Experiment 4 (2 excluded).

5.1.2. Stimuli and procedures

The beginning of each trial in Experiment 4 was identical to Experiment 1 except that 500 ms after the start of a trial, 4 identical white squares appeared (1-pixel black border, side length 130 pixels). The centers of the squares were as follows, relative to the screen center point: (−180, 275), (180, 275), (−180, −275), (180, −275). Each square then moved, with a sliding motion, at 180 pixels per second for about 1517 ms, so that they ended in the following positions (180, 180), (−180, 180), (180, −180), (−180, −180), thereby covering the 4 corners of the disc array.

From this point, the trials were again identical to those in Experiment 1. When disc motion ceased, the four white squares slid back to their starting positions, an arrow cursor appeared in the display, and participants used the mouse to click on the disc that they believed to be the target. We analyzed results from 2322 trials in Experiment 4.

5.2. Results

The occluders severely reduced performance for Translation compared to Rotation; $t(19) = 7.49$, $p = 2.18*10^{-7}$, d = 1.68. Fig. 5B and C plot these results. Moreover, tracking during Rotation remained robust (around 90% on average).

5.3. Discussion

Experiment 4 demonstrates that tracking rotation in a shape-centered reference frame is resilient to occlusion. The severe cost associated with Translation in this experiment further demonstrates that the initial stimulus is represented as a square with internally referenced tagging of target identity. In a standard MOT experiment, the linear motion of a target should present no problem for tracking procedures. But because those procedures operate in an external reference frame, here, they conflict with the initial representation of the target as a location within a square; that is, as one of the discs in a square comprising several, as opposed to one among several that are organized as a square. In this case the distinction carries severe consequences.

6. Experiment 5: Tracking a circle

When one thinks of rotation, one generally thinks of a circle. We employed squares in the experiments so far because if the discs were organized around a circle, translating and rotating motion would look identical. With the basic effect established, though, a circle

is an interesting stimulus to investigate. Consider how one might program the stimulus to understand why. Coding sixteen independent discs, each assigned a position at a fixed distance from the origin and at angles from 0° to 360° in 22.5° increments—this would be an externally referenced representation (using polar coordinates). One could then code a routine to serially update the position of each disc at some angular increment. Such a routine takes externally referenced coordinates as input and produces updated (and externally referenced) coordinates as output. And this is an implementation of *translation* in the sense that each disc is updated individually, albeit with the same routine applied to each. The discs remain independent objects in this approach. Indeed, nearly the same approach is what we used to program translation around a square perimeter in Experiment 1.

As an alternative, one could define a circle, making it virtual by setting its color to the background. Then one could define sixteen locations on the circle perimeter and draw a disc at each of the locations. Finally, one could use a rotation routine to rotate the circle. Rather than take the externally referenced coordinates for the discs as input, the routine would take the externally referenced coordinates of the circle as input (origin, radius length, and current orientation of the diameter axis), outputting an updated orientation for the diameter axis, and redrawing the whole circle accordingly, discs included. This implements *rotation* and it employs internal reference.

Interestingly, coding either way produces the same visual impression. Circular motion thereby admits both rotation and translation as interpretations. Which interpretation does perception rely on? We expect that such a stimulus is represented as a rotating circle, that the visual system reverse engineers the stimulus using a composite representation including *externally* referenced coordinates for the circle formed by the discs, *internally* referenced coordinates for the locations of the discs, and relational variables that characterize the angular tilt of the circle axis relative to external ones. If so, then it should be easier to track a target on a circle than a target on an equivalent square with discs translating around its perimeter, even though the stimuli include nearly identical visual motions (in terms of pixel-by-pixel movement). Experiment 5 tested this prediction (Fig. 6). A demonstration of the experiment can be viewed here.

The circle provides another useful feature for present purposes: unlike a square, it does not supply corners, sides, or other categorical landmarks to support descriptions such as “upper left” or “one item from the corner.” Effective tracking of a rotating circle would therefore provide evidence that the Rotation advantage does not depend entirely on categorical or verbal coding.

6.1. Methods

The registration for Experiment 5 can be found here.

6.1.1. Participants

21 participants were tested for Experiment 5 (1 excluded).

6.1.2. Stimuli and procedures

In Experiment 5, the Square/Translation condition was identical to the equivalent condition in Experiment 1. For the Circle/Rotation condition, we arranged the discs on the perimeter of a circle. The circle had a diameter equal to the diagonal of the square and rotated at the same angular speed as the Rotation condition in Experiment 1. Consequently, every target disc on the circle moved at a linear speed of 0.56 pixels/ms, tracing a circular path with a circumference of 1688 pixels. This was equivalent to the speed of a corner target in the square Rotation condition of Experiment 1. The linear speed in Translation trials was again 0.50 pixels/ms, meaning that targets in the Rotation condition moved approximately 12% faster than targets in the Translation condition. We analyzed results from 2386 trials in Experiment 5.

6.2. Results

Experiment 5 confirmed our prediction. Tracking performance was significantly better in Rotation (circle) trials than in Translation (square) trials on average, $t(19) = 2.86$, $p = 4.97{*}10^{-3}$, d = 0.64. Fig. 6B and C plot the results.

6.3. Discussion

In the previous experiments, the use of squares meant that participants could achieve something close to 25% accuracy, even without tracking per se. This is because a square allows for encoding the target as, for example “in the middle of one of the sides,” which could then support a guessing strategy that should perform far better than 1/16. Such a description is internally referenced—it refers to the structure of the square—but it would not necessarily be a perceptual representation. It could be entirely verbal. This is not a serious concern because performance in all the square experiments was far better than 1/16 for Rotation. Still, effective tracking on a rotating circle in Experiment 5 is strong evidence for a bona fide representation that includes an internal axis of reference. The only way to reidentify the target in this experiment is by reference to a structure not present in the raw input data—an internal axis—and by tracking the orientation of that structure relative to an external frame.

7. Experiment 6: Replication with a limited choice paradigm

Experiment 6 was designed to replicate once more the main finding of the previous experiments, by including the rotating square, the translating square, and the rotating circle condition in a single experiment. Additionally, we sought to further consider the possibility that square rotation benefits from a categorical representation of the stimulus, one that is not necessarily internally referenced

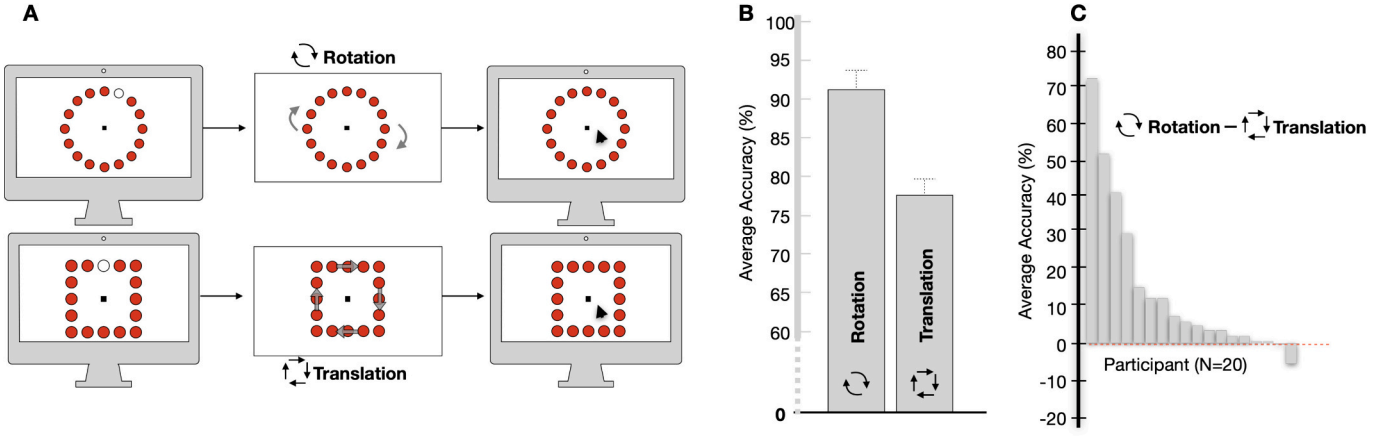


Fig. 6. (A) Stimuli and task (not to scale) for Experiment 5. The Translation condition is identical to Experiment 1. The Rotation condition employs a ring. (B) Average tracking performance for the two motion conditions. (C) Difference in average tracking performance (Rotation – Translation) for each of the 20 participants.

in a perceptual sense. Suppose a participant tracks their target by labeling it as in the middle of the square's top edge, for example. If the participant loses track of where the square's top sits currently relative to the external world—thereby losing track of the target—the participant could still avail themselves of an effective guessing strategy: choose a target in the middle of one of the four edges. This would allow for chance performance of 25%. As we noted in Experiment 5, performance in these experiments is well above 25% across the board. Nonetheless, we sought to determine if the advantage for square rotation might owe to efficient guessing at the margins by placing all conditions on common footing. Toward this end, we altered the procedure so that at the end of each trial, all but four of the red discs were removed from the test display, with the true target always among the set (Fig. 7A). Participants selected from among those four, offering a baseline guessing rate of 25% in square translation and circle trials, to correspond with the strategic guessing rate available in square rotation trials. We predicted that square rotation and circle rotation trials would nonetheless prove easier than square translation trials, eliminating strategic guessing at the margins as the cause of observed performance. A demonstration of the experiment can be viewed here.

7.1. Methods

The registration for Experiment 6 can be found here.

7.1.1. Participants

20 participants were tested for Experiment 6b.

7.1.2. Stimuli and procedures

The circle condition and the square translation condition were the same as Experiment 5. The square rotation condition was the same as in Experiment 1. At the end of each trial, all discs disappeared except for 4 discs in opposite positions, with the true target always included among them. Participants chose a response from this limited set. We analyzed results from 2962 trials in Experiment 6.

7.2. Results

A one-way repeated-measures ANOVA showed a significant performance difference among the three motion types, $F(2, 38) = 19.04$, $p = 1.87*10^{-6}$, $\eta_p^2 = 0.5$. Planned one-tailed paired *t*-tests showed significantly better tracking performance for the Rotation of a square than for Translation of a square, $t(19) = 4.97$, $p = 4.29*10^{-5}$, $d = 1.11$; better tracking performance for Rotation of a square than for Rotation of a circle, $t(19) = 3.05$, $p = 3.27*10^{-3}$, $d = 0.68$; and better tracking performance for Rotation of a circle than for Translation of a square, $t(19) = 3.90$, $p = 4.77*10^{-4}$, $d = 0.87$.

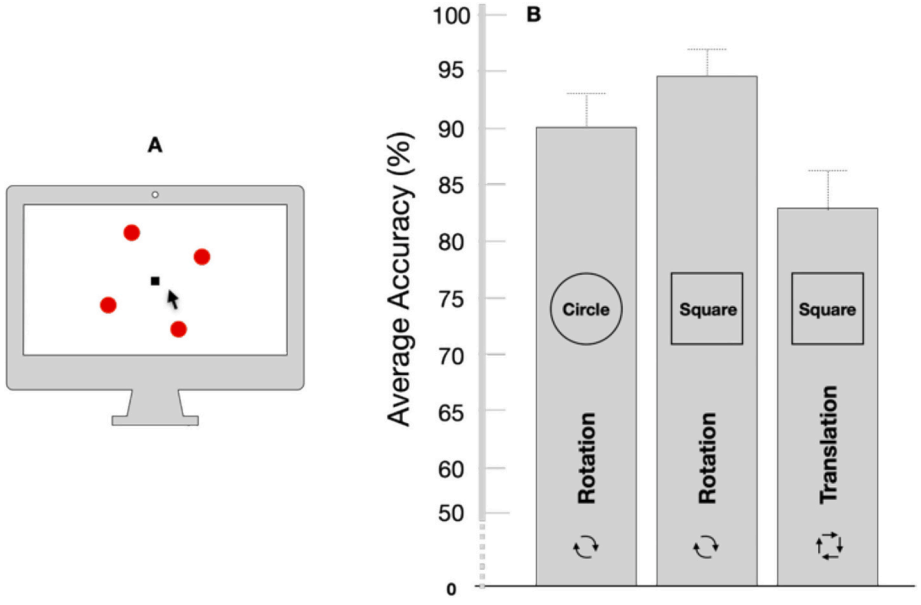


Fig. 7. (A) Test display (not to scale) for Experiment 6. The same test display was preceded by one of three trial types, square rotation (as in Experiment 1), square translation (Experiment 5) or circle rotation (Experiment 5). (B) Average tracking performance for the three motion conditions.

7.3. Discussion

Although participants had a 25% guessing probability in all three conditions, they performed better in both Rotation conditions than in the Translation condition. Better performance for the square rotation compared to the circle suggests that differentiation between positions on a square improves performance at the margins. This advantage could accrue because a square prevents target loss during rotational tracking or because it fosters recovery of lost targets. The latter seems unlikely to us because accurate recovery should occur at a rate of 25% at best. Thus, if targets become lost equally often in the square and circle rotation conditions, limited final selection in the current experiment should have fostered an equivalent level of performance in the Circle rotation condition. We therefore suspect that the difference owes to a difference in the frequency of losing a target. We speculate that this could be caused by mandatory representation of multiple internal axes in a square (which is a symmetric rectangle) versus only a single axis for a circle. Alternatively, because a rotating circle admits interpretation as a set of translating units, at least in principle, it may occasionally tend toward such a representation. But further experiments are required to draw definitive conclusions. On the whole, the experiment makes it unlikely that observed differences between square Rotation and Translation arise from a categorical or verbal coding strategy that is not properly internally referenced.

8. Experiment 7: Reducing internal reference through disaggregation

Finally, we reasoned that if we could devise a manipulation that somewhat disaggregates the discs in the square from one another, it would *improve* tracking performance under translation. Toward this end, in Experiment 7 we connected each disc by a one-pixel line to a partner disc reflected across the midpoint of the square (Fig. 8). A uniform and continuous connection will typically elicit a representation of parts as belonging to the same object (Behrmann & Tipper, 1994, 1999; Watson & Kramer, 1999). In the present case, we anticipated that this could lead to a representation of eight dumbbells organized so that their ends form the perimeter of a square. Whereas sixteen individual discs may be subsumed entirely under a representation of a well-organized square, dumbbells might suggest a second organization of the elements, and therefore they could lead more naturally to a dual representation: here is a square with spokes, and it might rotate, but it may also be a collection of dumbbells that will move in pairs. The experiment also included unattached discs, exactly as in Experiment 1. A demonstration can be viewed here.

8.1. Methods

The registration for Experiment 7 can be found here.

8.1.1. Participants

20 participants were tested in Experiment 7 (no exclusions).

8.1.2. Stimuli and procedures

Experiment 7 was identical to Experiment 1 except that in half of the trials a single-pixel black line connected pairs of discs that sat directly across from one another in the square arrangement. Each participant completed 240 trials of the experiment, with Translation/Rotation trials and connected/unconnected trials distributed equally and randomly through a session. We analyzed results from 4757 trials in Experiment 7.

8.2. Results

The results replicated Experiment 1: better tracking performance for Rotation compared to Translation for a square comprising unattached discs, $t(19) = 4.29$, $p = 1.98*10^{-4}$, $d = 0.96$. When the discs were attached in pairs, the difference between these conditions was still significant but greatly diminished: $t(19) = 2.51$, $p = 1.07*10^{-2}$, $d = 0.56$. Crucially, better tracking performance was observed for connected translating discs compared to unconnected translating discs, $t(19) = 3.02$, $p = 3.51*10^{-3}$, $d = 0.68$.

8.3. Discussion

Our interpretation of the results overall hangs on the concept of representational conflict: that the structure of a stimulus can cause observers to initiate representations that conflict with the tracking procedure that should be ultimately preferred given the motion that ensues. Representing a disc as part of a square conflicts with the need for externally referenced marking of the disc's position if the disc translates. In Experiment 7, we sought to create ambiguity around the initial representation of the stimulus. If the unconnected stimulus can be classified as either “a square comprising sixteen discs” or “sixteen independent discs arranged around a square,” the line connections in this experiment introduced a third option, “eight dumbbells each comprising two discs.” This third representational format is more amenable to tracking under translation because once the target disc is identified, a tracking procedure can mark the side of the relevant dumbbell that includes the target.

9. Discussion

The present experiments tested the hypothesis that cognitive conflict arises when the procedural demands of a task are misaligned

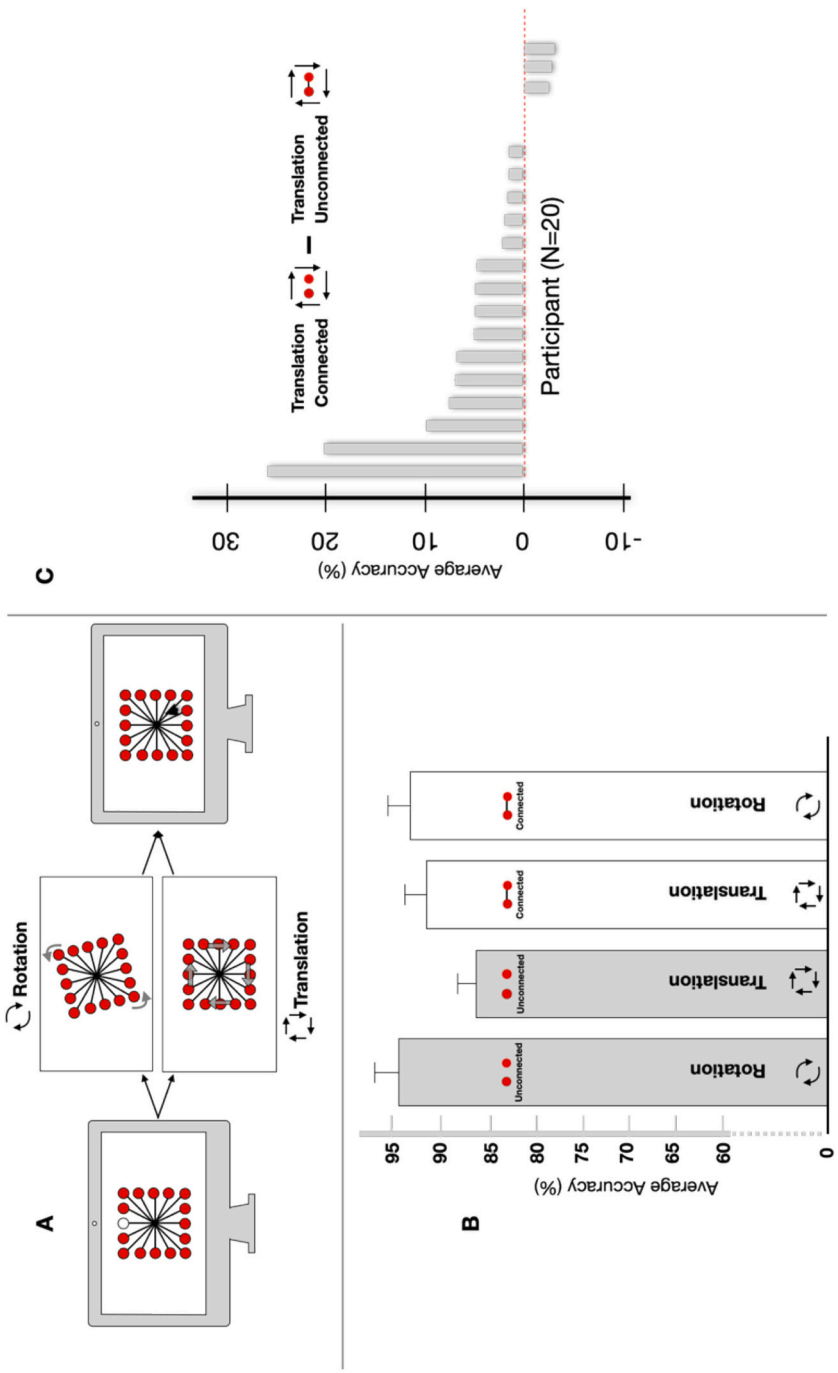


Fig. 8. (A) Stimuli and task (not to scale) for Connected trials in Experiment 7. Unconnected trials were identical to Experiment 1. (B) Average tracking performance by motion condition and connection type. (C) Participant differences in average tracking performance for Unconnected Translation – Connected Translation.

with the format of the internal representation naturally evoked by a stimulus. Specifically, we asked participants to track a target among identical nontargets. In Translation trials, where each item moved independently along a path, the most effective tracking procedure would utilize externally referenced coordinates. However, when items were organized into a coherent shape, such as a square, participants appeared to construct composite, internally referenced representations—representations that describe the parts in relation to an overarching structure. This made tracking Translation more difficult than in Rotation trials. An internally referenced representation aligns naturally with a rotation procedure. In contrast, translation requires independent tracking of individual elements, a procedure that is incompatible with the initial format of a shape-centered representation.

The conflict between format and procedural demand emerged reliably across seven experiments, and the performance costs under Translation were replicated and extended across stimulus types, including rounded-corner squares (Experiment 2), non-rectilinear perimeters (Experiments 3 and 5), and occluded displays (Experiment 4). Notably, these costs were attenuated in Experiment 7 when minimal structural disruption—via pairwise element connections—encouraged disaggregation of the composite shape, suggesting that representational format is malleable and sensitive to structure-inducing cues.

We have argued that rotation advantages performance in the experiments reported here by encouraging representation and tracking of the target as part of a larger whole. This may seem surprising in light of previous work showing that elements can become more difficult to track when they are represented as parts of a larger whole. Scholl et al. (2001) found that tracking multiple independently moving items becomes worse when targets and distractors are visibly connected into larger objects. The key aspect of that task, in our view, is that it demands externally referenced tracking of the targets and therefore conflicts with the internally referenced representations elicited by the stimulus. That and related studies are, in a sense, a mirror of the experiments reported here, where performance is best when motion and stimulus configuration agree on the underlying need for internally referenced representation. The lack of independence that often makes parts harder to track can become an advantage when the task can be solved in the coordinate system of the whole (see also Wu & Flombaum, 2024, Experiment 4, and Wu & Flombaum, 2026, Experiment 2).

A useful comparison comes from our recent work on the representation of objects in orbit ((Wu & Flombaum, 2026), which employed a closely related tracking paradigm but produced the opposite performance pattern. In those experiments, participants initially viewed sparse pairs of objects, with one member of each pair designated as a target (Fig. 9). Before motion onset, the sparse display was naturally compatible with representing the elements as independent objects, a format well suited to tracking their individual identities. The critical manipulation came when motion began. In translation trials, the objects continued to move as independent elements, preserving a representational format compatible with the tracking task. In rotation trials, however, the onset of orbital motion induced a reorganization: the same elements were now represented as parts of composite, orbit-centered systems. To see a demo of the previously published experiment, click here. Tracking was worse under rotation than translation. We argued that this deficit arose because motion reformatted the representation in a way that conflicted with the procedure already demanded by the task.

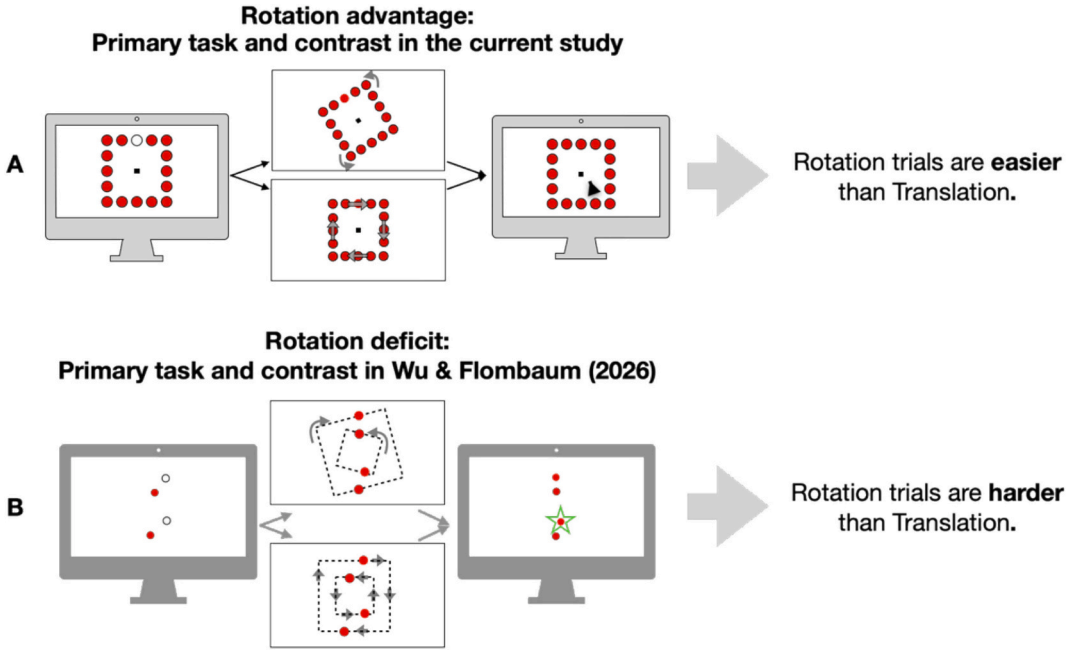


Fig. 9. Opposite tracking consequences of internally referenced representation across two paradigms. Top: In the present experiments, a target remains fixed within an internally referenced rotating structure, facilitating tracking relative to translation around the structure's perimeter. Bottom: In Wu & Flombaum, 2026, sparse orbiting targets are represented as parts of composite systems, making individual target identity more difficult to recover during rotation than during translation. Virtual squares are shown for illustrative purposes but were absent in the experiment. The opposite behavioral effects illustrate that internal reference is neither intrinsically beneficial nor detrimental. Performance depends on the compatibility between representational format and task procedure.

Individual targets became represented as parts of larger systems rather than as independently moving objects, so recovering their identities in externally referenced terms required additional computation. Thus, internally referenced representation does not make rotation intrinsically easier or harder. Its effect depends on whether the representational format aligns with the procedure required by the task.

The reversal also argues against generic low-level accounts of either set of experiments. Rotation and Translation in the present experiments necessarily differ in some local motion properties. In all the rotating-square conditions, we equated the time required for a complete revolution with the time required to complete a lap in the corresponding Translation condition. But this does not equate the instantaneous linear speed of every element, which depends on an item's position within a rotating square. However, performance did not vary meaningfully depending on whether the target occupied a midpoint, corner, or intermediate position on a rotating square, despite the corresponding differences in linear speed. These differences therefore do not appear to contribute appreciably to the observed performance effects. Moreover, Experiments 5 and 6 preserved the Rotation advantage for a rotating circle compared with translation around a square, even though targets in the Rotation condition actually moved faster than targets in the Translation condition. Abrupt changes in direction at square corners also do not appear to explain the effect. In Experiment 3, these direction changes occurred while the target was occluded; in Experiment 7, the same cornering was present even though connecting the elements improved performance. More importantly, Wu and Flombaum (2026) obtained the opposite effect using closely related rotational and translational motions. Although translational motion included comparable direction changes, translation facilitated rather than impaired tracking. Thus, rotation is not generically easier or harder because of speed, smoothness, path geometry, or some other fixed low-level property. Instead, the direction of the behavioral effect follows from the relation between representational format and task demands.

These findings offer insight into how perception constructs and utilizes internal representations. Internally referenced, object-centered formats are more than passive perceptual groupings. Rather, they constitute functional data structures that afford specific procedures (e.g., rotation) and resist others (e.g., tracking individual translation). This echoes principles in computation, where efficient performance depends on matching procedures to appropriate data formats. When such alignment fails, performance falters—even in tasks as simple as tracking a colored dot. Effective visual cognition—like computer programming—requires representations and procedures to comport with one another.

A potential implication of these results is an elaboration of the perceptual phenomenon of "grouping." A minimal definition from a comprehensive review states that grouping refers to "the fact that observers perceive some elements of the visual field as 'going together' more strongly than others (Wagemans et al., 2012, p. 1180)." If one were to assimilate the current results and interpretations into the concept of perceptual grouping, it could be classified further as a procedure that links together elements in an input image, minimally, by assigning them a shared (internal) axis of reference. It would follow that a group is a kind of representation, one that describes otherwise distinct elements with respect to shared axes that need not manifest physically in the raw stimulus. Further, a group would be a representation in the sense that it becomes the input for routines, such as tracking, with the form of the representation dictating the kinds of routines available to perform a task. To determine whether such an elaboration of the concept is warranted, future research could examine its implications in the context of the wider range of known grouping cues, such as common region and proximity.

These findings also extend prior neuropsychological research on object-centered representations, which has primarily drawn on evidence from individuals with perceptual deficits. Case studies of the individuals A.H. (McCloskey et al., 2006) and Davida (Vannuscorps et al., 2021) focus on systematic reflection errors that can only be explained in terms of representations that are structured relative to shape-intrinsic coordinates. Much of this earlier work relies on static stimuli and populations with atypical developmental or acquired deficits (e.g., Driver & Halligan, 1991). The present experiments demonstrate that object-centered representations are not limited to disordered cognition but are actively recruited in typical, unimpaired adults. Moreover, by employing dynamic motion stimuli we show that these representations are more than static encodings. They serve as scaffolds for perceiving dynamic changes in real time.

Finally, the results engage with foundational debates in the cognitive sciences. The performance costs observed under representational conflict are difficult to reconcile with direct perception theories that deny the role of internal representations. Instead, the evidence supports a view in which the visual system constructs abstract structural descriptions that go beyond the immediate sensory input. These descriptions include latent organizational features—axes, spatial relations, and shape-centric coordinates—that enable predictive and interpretive cognitive functions.

The findings from Experiment 5 highlight this point. Nearly identical visual motions (in terms of pixel-by-pixel movement) yielded distinct cognitive consequences depending on how the motion was construed—rotation of a circle versus translation around a square. That participants tracked more effectively when motion was interpreted as a rotation of a whole (rather than independent translations) supports the hypothesis that the visual system selects representational formats based on structural coherence. Perception, in this case, is not a mirror of motion, but an inferential process that organizes input into meaningful, manipulable formats.

Taken together, these findings reveal a general principle of cognition: that the compatibility between representational format and cognitive procedure is a hidden but powerful determinant of behavioral performance. Cognition—like software engineering—requires that representations and routines are aligned. When they are not, errors arise due to a fundamental mismatch between how information is represented and how it is meant to be used.

Studies in human

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place.

This study was approved by the Johns Hopkins University Homewood IRB.
(Approval No. HIRB00017514)

CRediT authorship contribution statement

Qihan Wu: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jonathan I. Flombaum:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization.

Informed consent and patient details

Written informed consent to take part in the study and to publish the article has been obtained from all participants or their legal representatives. The privacy rights of participants have been observed.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Behrmann, M., & Tipper, S. P. (1994). Object-based attentional mechanisms: Evidence from patients with unilateral neglect. *Psychologia, 25*(1A), 31–40.
Behrmann, M., & Tipper, S. P. (1999). Attention accesses multiple reference frames: Evidence from visual neglect. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 83.
Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems, 6*(1–2), 3–15.
Driver, J., & Halligan, P. W. (1991). Can visual neglect operate in object-centred co-ordinates? An affirmative single-case study. *Cognitive Neuropsychology, 8*(6), 475–496.
Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition, 107*(3), 904–931.
Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition*. Psychology press.
Hafri, A., Bonner, M. F., Landau, B., & Firestone, C. (2024). A phone in a basket looks like a knife in a cup: Role-filler independence in visual processing. *Open Mind, 8*, 766–794.
Ma, W. J., & Huang, W. (2009). No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *Journal of Vision, 9*(11), 1–30.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. WH Freeman and Company.
McCloskey, M., Valtonen, J., & Cohen Sherman, J. (2006). Representing orientation: A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology, 23*(5), 680–713.
Odic, D., Knowlton, T., Wellwood, A., Pietroski, P., Lidz, J., & Halberda, J. (2024). Observers efficiently extract the minimal and maximal element in perceptual magnitude sets: Evidence for a bipartite format. *Psychological Science, 35*(2), 162–174.
Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197.
Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences, 46*, Article e261.
Rokers, B., Yuille, A., & Liu, Z. (2006). The perceived motion of a stereokinetic stimulus. *Vision Research, 46*(15), 2375–2387.
Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*(2), 259–290.
Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145.
Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13*(2), 113–142.
Upadhyayula, A., & Flombaum, J. (2020). A model that adopts human fixations explains individual differences in multiple object tracking. *Cognition, 205*, Article 104418.
Vannuscorps, G., Galaburda, A., & Caramazza, A. (2021). The form of reference frames in vision: The case of intermediate shape-centered representations. *Neuropsychologia, 162*, Article 108053.
Vannuscorps, G., Galaburda, A., & Caramazza, A. (2022). Shape-centered representations of bounded regions of space mediate the perception of objects. *Cognitive Neuropsychology, 39*(1–2), 1–50.
Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin, 138*(6), 1172.
Watson, S. E., & Kramer, A. F. (1999). Object-based visual selective attention and perceptual organization. *Perception & Psychophysics, 61*(1), 31–49.
Wu, Q., & Flombaum, J. I. (2024). The motion-silencing illusion depends on object-centered representation. *Psychological Science, 35*(5), 504–516. https://doi.org/10.1177/09567976241235104
Wu, Qihan, & Flombaum, Jonathan I. (2026). The format of representation for objects in orbit: Composite and internally referenced. *Psychological Review*. https://doi.org/10.1037/rev0000629. https://doi.apa.org/doi/10.1037/rev0000629. (Accessed 18 May 2026)