

What Do Object Files Pick Out?

E. J. Green*†

Many authors have posited an “object file” system, which underlies perceptual selection and tracking of objects. Several have proposed that this system internalizes principles specifying what counts as an object and relies on them during tracking. Here I consider a popular view on which the object file system is tuned to entities that satisfy principles of three-dimensionality, cohesion, and boundedness. I argue that the evidence gathered in support of this view is consistent with a more permissive view on which object files select and track according to well-known perceptual organization criteria. Further evidence supplies positive support for the permissive view.

1. Introduction. Research suggests that vision deploys a system of object representations. This system has been called the *object file system*. Many have held that by virtue of internalizing certain principles about objects, the object file system is *tuned* to a specific type of entity. In this article, I consider the types of entities to which object files are tuned.

I focus on a view defended by Tyler Burge, Susan Carey, and others on which the object file system is tuned to *Spelke-objects*. Spelke-objects are characterized by several properties, including *three-dimensionality* (3-D), *cohesion*, and *boundedness*. I will call the view that the object file system is tuned to Spelke-objects the *restrictive view*. I will argue that we have little reason to accept the restrictive view.

In section 2, I distinguish two versions of the restrictive view, which I call the *strong restrictive view* and the *weak restrictive view*. The former view

Received September 2016; revised January 2017.

*To contact the author, please write to: Department of Linguistics and Philosophy, Massachusetts Institute of Technology; e-mail: ejgr@mit.edu.

†I am grateful to Frankie Egan, Jacob Feldman, Chris Hill, Brian McLaughlin, Jake Quilty-Dunn, and Susanna Schellenberg for helpful discussion and comments. Comments from two Philosophy of Science reviewers led to significant improvements. Thanks also to audiences at the 2015 European Society for Philosophy and Psychology; University of Nevada, Reno; University of California, Irvine; MIT, and Northwestern University.

Philosophy of Science, 85 (April 2018) pp. 177–200. 0031-8248/2018/8502-0001\$10.00
Copyright 2018 by the Philosophy of Science Association. All rights reserved.

claims that the object file system critically relies on representing 3-D, cohesion, and boundedness, while the latter view claims that object files have the proper function of picking out Spelke-objects. In section 3, I argue that current evidence fails to support the strong restrictive view over a competing position that I call the *permissive view*. According to the permissive view, the object file system tracks entities by attributing certain well-known perceptual organization cues. In section 4, I argue that if the strong restrictive view is false, then the weak restrictive view is likely false too.

2. The Restrictive Principles. Object representations play an important role in visual attention. Attention naturally “spreads throughout” an object, enabling quicker comparisons of features belonging to the same object than features belonging to different objects (Chen 2012). Moreover, multiple-object tracking (MOT) studies have shown that perceivers can attentively track a small set of objects as they move randomly amid distractors (Pylyshyn and Storm 1988), even as their features change unpredictably (Zhou et al. 2010).

Object representations are also used in visual working memory (VWM). First, object-reviewing studies have shown that we are faster to reidentify a previously seen feature when it reappears on the same object on which it initially appeared, even if the object shifts location in the interim (Kahneman, Treisman, and Gibbs 1992; Hollingworth and Rasmussen 2010). Second, change-detection studies have shown that it is easier to recall two features from the same object than from different objects (Luck and Vogel 1997; Olson and Jiang 2002). Thus, vision binds features to objects, and VWM capacity is set in part by the number of objects stored. There is also evidence that the representations used in VWM are continuous with those used in attentive tracking. For example, both MOT and VWM recruit the intraparietal sulcus (Drew et al. 2011), and objects held in VWM interfere with performance in MOT (Fougnie and Marois 2009).

Researchers have posited an “object file” system (OF-system) to account for these phenomena. The OF-system has three signature characteristics. First, object files sustain reference to entities over time, enabling us to track them from one moment to the next. Second, object files temporarily store features of the entities they pick out. Third, the OF-system is limited in how many entities it can select in parallel. On most views, the OF-system is limited to selecting three to four objects, with capacity slightly increasing between infancy and adulthood (Zosh and Feigenson 2009). Limits in this range have been observed for both MOT and VWM, again suggesting a common system for both capacities (Pylyshyn and Storm 1988; Luck and Vogel 1997).

Object files have played an important role in recent philosophical accounts of demonstrative reference (Murez and Recanati 2016). Some have even appealed to object files in explicating the basis of our capacity to think

about particulars (Pylyshyn 2007). For these projects, it is no doubt important to consider the basic question of which things object files pick out. For example, if demonstrative reference is fixed through visual attentional channels (Dickie 2010), it is important to know which things those channels deliver.

In several seminal papers, Spelke (1990, 1994) proposed that infants segment scenes in accordance with certain *object principles*. Following Spelke, subsequent authors—notably Burge and Carey—have proposed that OF-system processes (selection, tracking, and VWM) are governed by these principles. Burge writes: “To represent something as a body, the individual’s perceptual system must segment a three-dimensional whole from a surround. . . . Its doing so is governed by principles for identifying cohesiveness and boundedness of three-dimensional volume shapes. And it must be able to track the wholes over time, either in motion or at rest. Tracking depends on attribution of maintenance of cohesiveness and boundedness of volume shapes” (2010, 464). Burge thus proposes that when tracking particular objects (which he calls “bodies”), the visual system relies on the principles of 3-D, cohesion, and boundedness. He claims elsewhere that visual tracking abilities are “emphatically tied to bodies that maintain their cohesion” (461). Carey adopts a similar view: “Object files symbolize physical objects, by which I mean bounded, coherent, 3-D, separable, spatio-temporally continuous wholes” (2009, 97). Let us unpack what these principles mean.

The 3-D principle, on Burge’s construal, requires that objects are *volume-metric*, and thus are distinct from 2-D retinal regions and 2-D surface patches. Note, however, that while Burge and Carey endorse versions of the 3-D criterion, Spelke does not include it within her set of object principles.

As Burge and Carey acknowledge, the cohesion and boundedness principles are due to Spelke. Spelke’s rules state topological criteria. The cohesion principle is this:

Cohesion: “Two surface points lie on the same object only if the points are linked by a path of connected surface points” (Spelke 1990, 49).

Cohesion entails that objects are material, topologically continuous figures. For any two surface points x and y , if x and y belong to a single object, then y can be reached from x by following a continuous path of surface points. My cell phone satisfies the cohesion constraint because any two points on it are linked via a continuous path of surface points.

The boundedness principle is this:

Boundedness: “Two surface points lie on distinct objects only if no path of connected surface points links them” (Spelke 1990, 49).

Boundedness entails that objects are *maximal* connected figures. For any two surface points x and y , if there are distinct objects O and O^* such that x belongs to O and y belongs to O^* , then y cannot be reached from x by following a continuous path of surface points. My cell phone and toaster count as distinct objects under the boundedness principle, while two halves of my phone do not.

Some include *solidity* as another object principle. This principle states that distinct objects cannot physically interpenetrate. Looking-time studies suggest that infants are surprised by solidity violations (Spelke et al. 1992). However, others have questioned whether the OF-system really takes solidity to be a necessary condition on objecthood, since carefully constructed displays can elicit percepts of objects persisting through interpenetration (Scholl and Leslie 1999, 37–38). It is possible that we are surprised by solidity violations but do not always interpret them as disrupting object persistence. In any event, I will not consider solidity in detail here.

By the *restrictive view*, I mean the view that the principles governing visual object selection and tracking include the principles of 3-D, cohesion, and boundedness. I will refer to these as the “restrictive principles,” and I will call the entities that satisfy them “Spelke-objects.”

We can distinguish two versions of the restrictive view, which differ in the role they assign the restrictive principles. The *strong* restrictive view holds that in order to select and track an object, the OF-system must represent it as 3-D, cohesive, and bounded. This view is suggested by Burge’s claim that “tracking depends on attribution of maintenance of cohesiveness and boundedness of volume shapes” (2010, 464). The *weak* restrictive view holds that the restrictive principles characterize the type of thing that the OF-system has the *proper function* of selecting. On this view, the OF-system is tuned to Spelke-objects, but it need not attribute the properties constitutive of being a Spelke-object during tracking. I will argue that we have little reason to accept either version of the restrictive view.

3. The Strong Restrictive View. This section considers the strong restrictive view. I first contrast this view with a competing view of the properties relied on during tracking, which I call the *permissive view*. Sections 3.2–3.4 consider the three restrictive principles. In each case, I will argue (i) that the evidence garnered in support of the principle can also be explained by the permissive view, (ii) that there are experimental tests that may help to settle the dispute between these positions, and (iii) that there already exists evidence against the principle.

3.1. The Permissive View. Kimchi (2009, 25) characterizes “visual objects” as “elements in the visual scene organized by Gestalt factors into a coherent unit” (see also Brovold and Grush 2012; Chen 2012). The idea is that

the principles guiding visual object individuation are simply the rules of perceptual organization. These rules characterize the properties that selection and tracking processes represent and rely on. I believe that this idea has received insufficient attention in the object file literature.

We can divide perceptual organization principles into *grouping* principles and *parsing* principles. The former specify rules for composing smaller entities into larger entities, while the latter specify rules for decomposing larger entities into smaller entities. By “entities,” I will be as noncommittal as possible. Bodies, shadows, parts, and arbitrary mereological sums all count as entities. When I speak of “units,” I mean only those types of entities that can be singled out through perceptual grouping or parsing.

Traditional grouping principles include proximity, similarity, good continuation, and common fate (Wagemans et al. 2012). For example, the proximity principle states that entities that are close together tend to be grouped into units, while the common fate principle states that entities that move along similar motion paths tend to be grouped. Newer grouping principles include element connectedness and uniform connectedness (Palmer and Rock 1994). The former states that connected entities tend to be grouped. The latter states that regions of the visual field that have some uniform property (e.g., color) tend to be treated as units. Examples are shown in figure 1.

Two important parsing principles are the *minima rule* (Hoffman and Richards 1984) and the *shortcut rule* (Singh, Seyranian, and Hoffman 1999). The minima rule states that the boundaries between parts of an object tend to be found at negative curvature minima—places at which the contour of a shape is most concave. The shortcut rule states that part divisions tend to be made by linking curvature minima along the shortest paths possible. The

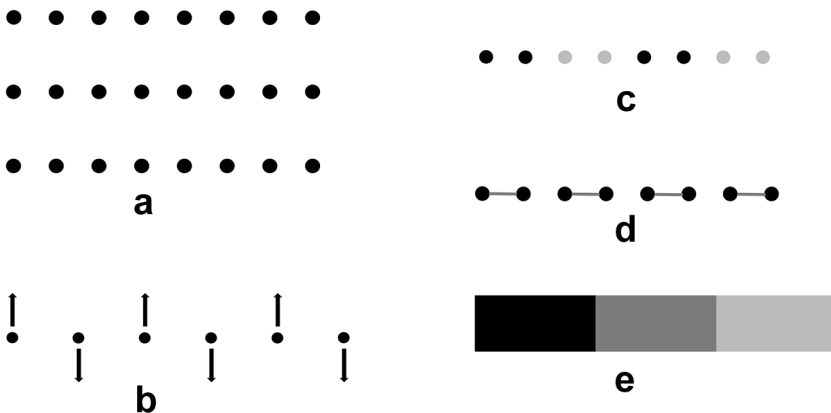


Figure 1. Grouping by (a) proximity, (b) common fate, (c) similarity, (d) element connectedness, and (e) uniform connectedness.

decomposition in figures 2*a* and 2*b* follows the minima and shortcut rules. There is strong evidence that vision segments objects in this way (Singh and Hoffman 2001) and that infants and young children engage in parsing (Tversky 1989; Bhatt et al. 2010). I call parts of objects returned by parsing principles *parsable parts*.

Grouping and parsing principles operate in natural scenes. Flocks of geese and swarms of bees seem like they can be differentiated and tracked over time. Likewise, it seems as though we can differentiate a person's arm and track it independently of her torso. But does the OF-system select and track entities such as these?

If so, this raises *prima facie* difficulties for the strong restrictive view. If the OF-system tracks groups while representing their noncohesion, then tracking does not critically rely on representing cohesion. If the OF-system tracks parsable parts while representing their nonboundedness, then tracking does not critically rely on representing boundedness. Note also that perceptual organization principles apply to both volumetric and 2-D entities. A group of 2-D dots can be organized by proximity, similarity, or common fate. If the OF-system tracks such units while representing their planarity, then tracking does not critically rely on representing 3-D.

By the *permissive view*, I mean, roughly, the view that the OF-system selects and tracks in accordance with perceptual organization principles and, moreover, that such principles characterize the properties represented and relied on during selection and tracking. The permissive view claims that when the OF-system selects and tracks an entity, it does so because it represents the entity as satisfying certain perceptual organization cues.

Multiple versions of the permissive view are possible, depending on which perceptual organization principles the OF-system follows. If the OF-system follows proximity, but not similarity, then it can treat a group of entities as a single object when the entities are tightly clustered but not when they

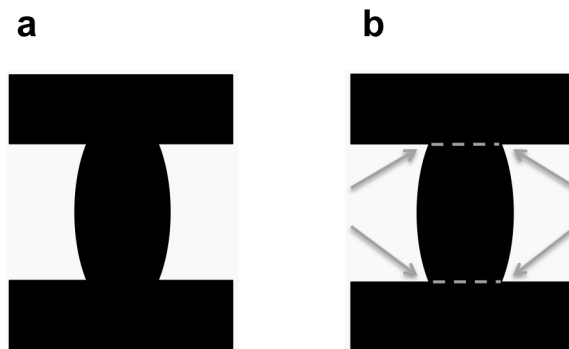


Figure 2. Decompositions following (a) minima and (b) shortcut rules.

merely share the same color. In what follows, I have in mind a specific variant of the permissive view. This is the view that, at all ages, object files can select entities either by representing them as satisfying common fate and proximity grouping cues or by representing them as satisfying minima and shortcut parsing cues. This view predicts that in certain situations, attribution of common fate and proximity is sufficient for the OF-system to select an entity, even if it is not represented as 3-D or cohesive. The view also predicts that in certain situations, the OF-system can select parts represented as satisfying the minima and shortcut rules, even if they are not represented as bounded.

Four clarifications are in order. First, the permissive view is consistent with internal connectedness (which figures prominently in the restrictive principles) being a cue to grouping and to tracking. However, unlike the strong restrictive view, the permissive view does not hold that attribution of internal connectedness is required for tracking. It instead places connectedness on a par with the other grouping cues.

Second, although I emphasize proximity and common fate, I believe it is plausible that at later ages, object files select entities on the basis of the other cues as well. There is evidence that certain grouping principles are acquired later than others, through learning or maturation (Quinn and Bhatt 2005). Moreover, below I discuss evidence that adults can select groups defined by similarity or good continuation.

Third, the permissive view does not entail that object files always select entities on the basis of geometrical simplicity. Spelke (1990) persuasively argues that infants do not automatically organize static scenes by contour alignment or symmetric form. This suggests that during infancy, the OF-system does not follow all of the perceptual organization principles. However, one can accept this while maintaining that infants follow certain organization principles. The common fate and proximity principles do not prioritize relations like symmetry or collinearity.

Fourth, I will not suggest that Spelke-objects, groups, and parts are treated alike by all visual processes. The present issue concerns only whether object files and the processes they subserves are keyed to the former category alone. Other visual processes may be different. I now turn to the evidence for the strong restrictive view.

3.2. Evidence for 3-D? Many of the tasks used to study object files employ 2-D images. It is also plausible from everyday experience that we can track certain nonvolumetric entities, such as shadows. Prima facie, these observations challenge the strong restrictive view. How can it be that the OF-system critically relies on attributing 3-D, if it so easily tracks 2-D entities?

The obvious response is that such 2-D entities are misrepresented as volumetric. Carey writes:

Does the fact that 2-D bounded entities activate object-files mean that their content is more perceptual—perhaps *closed shape*? . . . No, they should not. For computer displays to work, we must present many of the cues for depth in 2-D arrays, and surfaces arranged in 3-D are routinely perceived in such displays. That the system can be fooled into accepting 2-D entities as objects does not mean that it is not representing the stimuli as real objects, just as the fact that the system can be fooled into seeing depth in 2-D displays . . . does not mean it is not representing the stimuli as arrayed in 3-D space. (2009, 98)

Carey proposes, then, that computer images in tracking studies are misrepresented as “real” objects with volume. The evidence for this is that 2-D figures are selected only when they supply appropriate cues to depth.

The requirement that objects are “3-D” is somewhat ambiguous. On a stronger reading—which I have assumed so far—this means that objects have volume. But on a weaker reading, it just means that objects are arrayed in 3-D space, standing in depth relations to the observer and to other things. Something can have the second characteristic without the first. Surfaces are generally treated as 2-D entities arrayed in depth.

Even granting that 2-D computer images are represented as located in depth, this does not entail that they are represented as volumetric. So Carey’s response at best rescues the weaker version of the 3-D principle, not the stronger one. As we saw above, Burge understands the 3-D requirement in the stronger way. Carey is less explicit, but she may have in mind the weaker version. However, the weak reading seems quite forced. It counts planar polygons as “3-D,” as long as they are embedded in 3-D space. It should be noted, though, that the 3-D principle may be less important to Carey than the other principles, since her experiments do not generally test it.

Turning to the stronger version of the 3-D principle, there is very little evidence that the OF-system relies on representing tracked entities as volumetric. Stimuli in MOT and object-reviewing studies do supply cues to depth and occlusion (see Scholl and Pylyshyn 1999), and MOT performance is better when targets occupy a different depth plane from distractors (Rehman et al. 2015). However, such stimuli rarely supply standard cues to volume (e.g., differential shading, texture density, or surface orientation edges). Moreover, such stimuli look nonvolumetric. They look like flat, 2-D figures. As such, there is little evidential support for the claim that tracking relies on attributing volume.

To make progress on this issue, future work should systematically examine object selection and tracking as a function of cues to planarity versus volume. While I have suggested that they look 2-D, the stimuli generally used in midlevel vision tasks are arguably consistent with either a 2-D or 3-D interpretation.

There is, however, suggestive evidence that we can visually select entities that are quite clearly 2-D. De-Wit, Milner, and Kentridge (2012) found that shadows elicit typical object-based attention effects. When cued to a region of a cast shadow, subjects are faster to shift attention to another region of the same shadow than to an equidistant region of another shadow. Furthermore, children can detect shadows and acquire information from them at an early age (Cameron and Gallup 1988). It remains to be seen whether shadows can be selected in other OF-system paradigms.

3.3. Evidence for Cohesion? Studies of infant object cognition and adult midlevel vision have been taken to show that the OF-system relies on representing cohesion. I will argue that the available evidence does not in fact show this, since it can be explained equally well on the hypothesis that the OF-system selects and tracks by representing proximity and common fate cues. I will also discuss positive evidence that object files can select noncohesive entities.

Huntley-Fenner, Carey, and Solimando (2002) contrasted 8-month-olds' ability to track cohesive objects with their ability to track noncohesive sand piles. Infants in the 'object' condition saw a rigid entity that had the same shape, color, and texture as a sand pile. The object was lowered onto a stage while maintaining cohesion. Two screens were then introduced, one of which occluded the object. Next, infants saw another pile-shaped cohesive object lowered behind the second screen. Finally, both screens were removed, revealing either one or two objects. The 'sand' condition was identical to the object condition, except that the experimenter poured loose sand out of a cup into piles on the stage, rather than lowering cohesive objects.

In the object condition, infants looked longer toward the unexpected outcome of only one object, while in the sand condition, there was no significant difference in looking times for the two outcomes. This is consistent with the proposal that object files were maintained for cohesive objects (enabling infants to track how many there were) but were discarded for noncohesive sand piles.¹ The proposal is that once the piles were perceived as noncohesive, the OF-system could no longer track them.

Similarly, vanMarle and Scholl (2003) compared the standard MOT condition with a condition in which each object disintegrated into small pieces and "poured" from one location to the next. Tracking was significantly worse in the latter case (89% vs. 67% accuracy). More recently, Cheries et al. (2008) showed that simply splitting an object into two pieces disrupts infants' object representations. Infants were divided into a 'no-split' condition and a 'split'

1. These results should not be taken, however, to demonstrate that infants are wholly incapable of noticing changes in the quantity of a substance such as sand or water (Rips and Hespous 2015).

condition. Those in the no-split condition saw one cracker placed in a container and two already disconnected crackers placed in another container. In the split condition, they also saw one cracker placed in one container and two crackers placed in the other container. However, the two crackers resulted from breaking a single larger cracker within view of the infants. Cheries et al. found that infants reliably crawled toward the two-cracker container in the no-split condition, but were at chance in the split condition. They suggest that infants represented the loss of cohesion from splitting and thus discarded the object file for the larger cracker. Furthermore, infants did not have enough time to assign new files to the resulting pair of crackers before they were placed in the container.²

However, there is reason to suspect that cohesion violations may not be the key factor underlying tracking failures in these studies. This is because (i) failures in these or similar paradigms are observed with cohesive stimuli, and (ii) successes in these or similar paradigms are observed with non-cohesive stimuli.

As regards i, tracking is also impaired when entities retain their cohesion but expand or contract the way a pile of sand changes shape when poured from a cup. For instance, vanMarle and Scholl (2003) found that when targets expanded and contracted like “slinkies” (while maintaining cohesion), tracking was just as impaired as when they poured from one location to another (see also Howe et al. 2013). This raises the possibility that the tracking impairments in both Huntley-Fenner et al. (2002) and the pouring condition of vanMarle and Scholl (2003) may be due to the expansion and contraction involved in pouring, rather than loss of cohesion per se.

As regards ii, there is evidence that tracking is unimpaired for non-cohesive groups, as long as entities in a group move as a tight cluster. In a fourth condition of vanMarle and Scholl’s (2003) study, noncohesive groups of items that moved as clusters were tracked just as well as standard figures in the MOT task. Similarly, Wynn, Bloom, and Chiang (2002) found that infants are capable of enumerating perceptual groups, independently of the number of elements in those groups. Infants were habituated to a display containing either two or four moving clusters of three dots. During test trials, they saw displays containing either two groups of four dots or four groups of two dots. Infants who habituated to displays containing two groups looked longer at test displays of four groups, while those habituated to displays containing four groups looked longer at test displays of two groups. Note that both test displays contained the same number of individual dots. A natural explanation, then, is that infants assigned files to each group of dots on the

2. These findings mirror work on adult midlevel vision (Mitroff, Scholl, and Wynn 2004). Note, however, that Cacchione (2013) found that infants could succeed in ‘split’ conditions when the larger (presplit) graham cracker was four times as large as the smaller cracker, or when it had a center piece removed, rather than being split in half.

basis of common fate and proximity, enabling them to detect changes in the number of groups.³

The foregoing considerations suggest a competing explanation of the results taken to support cohesion. The reason for failures of tracking and individuation in these cases may be that the OF-system is highly sensitive to correlated motion paths among clustered items. On this hypothesis, items are grouped by representing cues to proximity and correlated motion. However, given cues to independent motion among items, items are unlikely to be grouped—or, if they are already grouped, they are likely to become ungrouped. The attribution of proximity and common fate is, on this proposal, sufficient for selection and tracking. Further attribution of cohesion is unnecessary.

Consider again vanMarle and Scholl's (2003) slinky condition. The present hypothesis suggests that although the entities maintained cohesion, the visual system detected that the elements comprising them (their front and back edges) moved independently. This could have prevented grouping of the edges into a single object and led to a failure to maintain object files for the slinkies. This proposal also applies to the "pouring" condition of vanMarle and Scholl's study and to the Huntley-Fenner et al. (2002) task, since the elements of the poured substances provided strong cues to independent motion. Moreover, consistent with the proposal that tracking relies on correlated motion cues, St. Clair, Huff, and Seiffert (2010) found that when texture elements within a target move opposite its motion direction, MOT performance suffers.

A similar hypothesis may explain the data from Cheries et al. (2008). When the cracker split apart, the resulting pieces initially followed quite different trajectories. The current hypothesis suggests that because of this, they were no longer perceptually grouped. Again, on this proposal it is not the violation of cohesion that explains tracking failures but, rather, cues to independent motion.

Likewise, in studies demonstrating tracking or individuation success despite lack of cohesion (Wynn et al. 2002; vanMarle and Scholl 2003), the key feature, on this proposal, is that the elements of noncohesive groups followed similar motion paths. This led them to be grouped into a unit and targeted by an object file.

3. Wynn et al. maintain that it is unlikely that groups were treated as single objects, because the dots within a group followed slightly different motion trajectories. However, this inference is questionable, because while dots in a group did not follow precisely parallel trajectories, their velocities plausibly were sufficiently correlated for common fate grouping. There is evidence that the neural pooling of motion signals is a flexible process that accommodates variance in the speed and direction of motion among a collection of items (e.g., Webb, Ledgeway, and Rocchi 2011).

Might the tracked groups in Wynn et al. (2002) and vanMarle and Scholl (2003) have been represented as cohesive? A defender of the strong restrictive view might appeal to some of Spelke's earlier work to motivate this proposal. Spelke and colleagues demonstrated that when infants encounter surface fragments that move together on opposite ends of an occluder, they perceive them as belonging to a single, cohesive object. For example, if infants see the ends of a rod move back and forth behind a box, then they will subsequently look longer toward a pair of fragmented rods than toward a single, connected rod (Kellman and Spelke 1983; Kellman, Spelke, and Short 1986).

Kellman and Spelke concluded that from an early age, infants can compute cohesion from common motion. One might suggest that this occurred in Wynn et al. (2002) and vanMarle and Scholl (2003). However, in Kellman and Spelke's paradigm, an occluder separated the moving fragments. Accordingly, infants were not provided with any evidence that contradicted a cohesive interpretation. Thus, this evidence does not show that vision infers cohesion from common motion even when contradictory information is available—namely, when discontinuities between fragments are fully visible. The stimuli in Wynn et al. (2002) and vanMarle and Scholl (2003) fall into the second category. Thus, to motivate this response the defender of the strong restrictive view would need to show that participants in these studies represented cohesion despite inconsistent visual information.

Future research should directly investigate whether infants' apparent sensitivity to cohesion is really due to common fate or proximity grouping. Are infants also impaired when required to individuate and track "slinky" items, just as adults were in vanMarle and Scholl (2003)? If so, then common fate likely explains the Huntley-Fenner et al. (2002) results. Furthermore, it is critical to determine whether the clusters of elements tracked in Wynn et al. (2002) and vanMarle and Scholl (2003) were represented as cohesive, despite visible spatial gaps. Do participants perceptually distinguish moving clusters from Spelke-objects but track them anyway? If so, then this is compelling evidence for the permissive view. If not, then the strong restrictive view may yet handle our capacity to track groups.

There is evidence beyond MOT, however, that object files may target noncohesive groups. Like Spelke-objects, perceptual groups structure the storage of information in VWM. Participants are better at recalling features from two grouped items than from two ungrouped items (Woodman, Vecera, and Luck 2003; Peterson and Berryhill 2013). Moreover, Brady and Tenenbaum (2013) recently showed that VWM capacity for patterned, many-item displays could be modeled reasonably well on the assumption participants "chunk" items based on similarity and proximity grouping cues. They estimated a capacity of four groups, mirroring limits generally found for cohe-

sive objects.⁴ Finally, there is evidence that groups elicit typical object-based attention effects. Feldman (2007) found that participants were faster to compare probes that appeared on line segments grouped by good continuation, symmetry, or parallelism than probes that appeared on ungrouped line segments.

3.4. Evidence for Boundedness? The boundedness requirement precludes undetached parts from counting as objects. But is there evidence that the OF-system relies on attributing boundedness? If so, then infants and adults should display pronounced deficits when asked to track entities seen as nonbounded. I will argue that while the evidence suggests that object files cannot target *arbitrary* parts of objects, it does not show that object files cannot target *parsable* parts of objects. I will then argue that further evidence suggests that object files can target parsable parts.

Although boundedness has received less attention than cohesion, several authors have suggested that the vision fails to track undetached parts. For example, Fodor and Pylyshyn (2015) write: “It turns out that subjects can track dumbbells but can’t track their weights (unless we remove the rod that connects them). Connecting the parts of the dumbbell creates a single new object, not just an arrangement of the parts of one (see Scholl, Pylyshyn, and Feldman 2001). . . . What subjects see when they see dumbbells is not *arrangements of undetached dumbbell parts* but, unsurprisingly, dumbbells. The world is the totality of *things*, not undetached parts of things” (131). Fodor and Pylyshyn suggest that the cited experiment (Scholl et al. 2001) shows that tracking is selectively keyed to bounded things, rather than to undetached parts.⁵ They also suggest that this helps in responding to Quine’s “gavagai” problem. While I will not be concerned here with referential indeterminacy issues, it is worth asking whether the evidence that they cite actually supports the claim that tracking mechanisms incorporate a strict boundedness constraint.

Fodor and Pylyshyn claim that subjects cannot track the weights of dumbbells unless we remove the rod that connects them. But the study they cite does not substantiate this. Scholl et al. (2001) examined a number of tracking

4. More precisely, Brady and Tenenbaum (2013) found that performance in such displays was best explained by a hybrid model on which VWM sometimes uses this chunking strategy and sometimes represents pattern statistics of the scene, such as how smoothly textured it is, along with four individual outliers.

5. Fodor and Pylyshyn may endorse a more general claim—namely, that in addition to not tracking undetached parts, we do not even “see” or visually differentiate undetached parts. However, this interpretive issue does not seem to affect the arguments below. For if we do not even differentiate undetached parts, then presumably we should also be incapable of selectively tracking them.

conditions, but three are most relevant. The first condition replicated the standard MOT paradigm. In the second, subjects tracked the endpoints of lines as they moved about in the presence of distractors. In the third, subjects tracked the square ends of dumbbells. Dumbbells were composed of two squares linked by a line segment. Importantly, dumbbell weights are likely to be discriminated through perceptual parsing, while the endpoints of lines are not—endpoints are arbitrary parts. So if the OF-system can select parsable but not arbitrary parts, then it should be able to select dumbbell weights but not endpoints.

Consistent with Fodor and Pylyshyn's claim, subjects were better in the standard MOT condition than in either of the other two, and performance in the third condition (tracking endpoints) was only slightly better than chance (see also Howe, Incledon, and Little 2012). However, performance in the second condition (tracking weights) was well above chance. Tracking accuracy with weights was roughly 84% versus 92% in the standard condition (Scholl et al. 2001, 170). Thus, while there was some (statistically significant) decrement associated with tracking weights versus tracking individual boxes, this hardly shows that we cannot track them.

There is also an unfortunate asymmetry in comparing tracking of "whole" objects versus parts of objects. Suppose that object files can select both bounded figures and parsable parts of them. Further, suppose that during tracking, we sometimes mix up a target with one of its parts. In this case, then we would expect MOT performance to be slightly better for whole objects than for parts. Tracking a part of an object is sufficient for reidentifying the whole object at the end of a MOT trial, while tracking a whole object is insufficient for reidentifying one of its parts at the end of a MOT trial. Thus, if a participant needs to track a whole object but "swaps" it with one of its parts, she will still perform the task equally well. However, if she needs to track only a part of an object but swaps it with the whole object (or an adjoining part), her performance will suffer. Thus, even if tracking accuracy is somewhat higher for whole objects than for parts, we cannot be sure whether this reflects a systematic deficit in selecting nonbounded entities or occasional object-part swapping.

It is important, then, to consider evidence beyond MOT. In another study taken to support boundedness, Mitroff, Scholl, and Wynn (2005) analyzed the effect of merging on the object-specific preview benefit (OSPB). Participants saw an initial display containing three circles. Letters briefly appeared on each circle and then vanished. Next, the objects took part in one of two events. Either two of the circles merged to become a single circle, or two of the circles approached one another without merging. Finally, a single letter appeared in one of the circles, and subjects had to indicate whether it had appeared earlier. OSPBs are associated with faster "same" responses to letters that reappear on the same object than to letters that reappear on a dif-

ferent object. Mitroff et al. found that merging (but not approaching) eliminated OSPBs for the lower of the two merged objects. This demonstrates that object files treat the merging of two circles into one circle as a transition from two objects to one. But why? One explanation for this is that the OF-system strictly relies on boundedness in determining object persistence. When the merged objects were no longer perceived as bounded, the OF-system concluded that one of them went out of existence and discarded the relevant object file.

However, we need not conclude that the OF-system relies critically on attributing boundedness in order to explain this result, because perceptual parsing rules would also require treating the merged circle as only one object. Since a circle lacks curvature discontinuities, there is no basis for perceptually decomposing it into parts. Thus, even if the OF-system merely relies on grouping and parsing cues, rather than boundedness per se, it should arguably treat the premerge circles as two objects and the postmerge circle as only one.

To make progress on this issue, future work should systematically compare OSPBs for bounded objects, arbitrary parts, and parsable parts. Are OSPBs also found when a preview feature reappears in the same parsable part of an object versus a different one? Moreover, it is critical to determine whether boundedness violations disrupt perception of object persistence in the context of preserved parsability. For example, suppose that two objects merge, but their merging preserves curvature minima that allow the figure to be decomposed into parts (see fig. 3). The strong restrictive view predicts that at least one of the OSPBs should be eliminated in this case, since the resulting entities violate boundedness. If, however, OSPBs are preserved through such transformations, this would support the permissive view.

Although this question awaits further investigation, there is already evidence to suggest that object files can target parts. First, like Spelke-objects and groups, parsable parts structure information storage in VWM. Two fea-

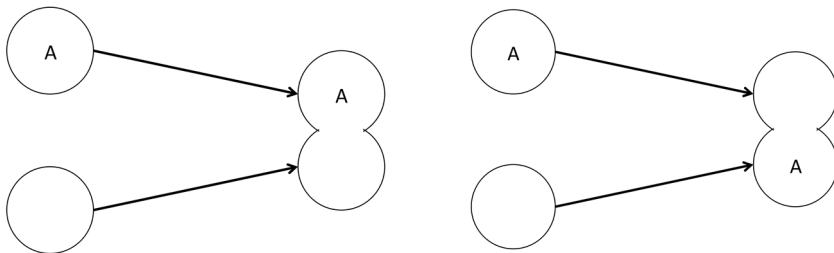


Figure 3. Merging with preserved parsability. *Left*, congruent match; *right*, incongruent match. OSPBs would be indicated by faster reaction times on congruent versus incongruent trials.

tures are easier to recall when they belong to the same part of an object than when they belong to different parts (Xu 2002). Moreover, visual attention more readily spreads within a part than across parts (Barenholtz and Feldman 2003), mirroring typical object-based attention effects. Finally, Porter et al. (2016) recently found that parallel enumeration, or “subitizing,” is similarly efficient for parsable parts (protrusions from a disc) and bounded objects. Enumeration time increased very gradually up to four parts (a mark of parallel selection), after which the slope steepened. This matches the OF-system’s set-size limit, and it is commonly held that subitizing recruits the OF-system (Pylyshyn 2007; Carey 2009, chap. 4).

3.5. Nonmaterial Things. I have argued that the evidence for the strong restrictive view is inadequate because it can also be explained by the permissive view. The strong restrictive view also faces another challenge. Note that the cohesion and boundedness principles apply to entities composed of surface points. This suggests that object files only target material things. But evidence suggests that this may be incorrect.

Holes are not composed of material surfaces but instead of empty space wholly surrounded by material surfaces. Nonetheless, holes can be tracked just as efficiently as standard stimuli in the MOT paradigm (Horowitz and Kuzmova 2011). Moreover, while perceivers are poor at remembering the shapes of arbitrary background regions, they remember the shapes of holes just as well as the shapes of solid objects (Palmer et al. 2008; although see Bertamini 2006). There is also evidence that we can track cavities or indentations in objects. Giralt and Bloom (2000) found no significant differences between 3-year-olds’ ability to track and enumerate cavities and their ability to track and enumerate solid objects. Furthermore, indentations can be rapidly individuated in parallel. In addition to parts, Porter et al. (2016) found typical subitizing effects for indentations in an object, suggesting similar set-size limits for indentations and Spelke-objects.

One might contend that nonmaterial entities can only be tracked if they are misrepresented as Spelke-objects. However, this is doubtful, since several studies of nonmaterial entities have provided ample information about surface depth. Giralt and Bloom used full-cue physical stimuli, while Horowitz and Kuzmova used holes defined by stereo disparity. Nonetheless, it remains to be seen whether nonmaterial entities elicit typical results in other paradigms, such as object reviewing.

3.6. Multiple Systems? Thus far, I have assumed a unified theory of tracking mechanisms. Specifically, I have assumed that if we can perform MOT or VWM tasks with both Spelke-objects and non-Spelke-objects, then object files are used in both cases. One might suggest, however, that vision

uses multiple systems for selection and tracking. Perhaps the OF-system is just one of them, and it follows the restrictive principles.⁶

There are *prima facie* reasons to doubt this multisystem account. To make the case that separate systems are responsible for selecting Spelke-objects and non-Spelke-objects, one needs to show that performance differs systematically between these stimulus categories. But this is precisely the assumption I have been questioning. For instance, there is little evidence for substantially different set-size limits across the various cases I have discussed. A multisystem account would render this a coincidence. Thus, if performance on standard OF-system tasks is similar for Spelke-objects and non-Spelke-objects, then this already damages the case for multiple systems.

Alternatively, one might suggest that the tasks I have discussed do not all recruit object files. Perhaps object files are engaged for VWM tasks but not for MOT, rendering MOT evidence irrelevant.⁷ As motivation for this, one might note that we often fail to remember properties of tracked objects during MOT (Bahrami 2003). This is puzzling, since object files are supposed to store properties of the objects they select. However, although participants are poor at recalling properties during MOT, this may be because MOT is so resource demanding that it often prevents properties from being entered into files (Pylyshyn 2007, 39–40). Tracking may still recruit object files, even if no properties are entered into them.

In any case, I have discussed evidence that VWM selects non-Spelke-objects, so the strong restrictive view's difficulties are not limited to MOT. Further, recall that there is evidence that common mechanisms are involved in tracking and VWM (Fougnie and Marois 2009; Drew et al. 2011). Likewise, tracking an object in MOT boosts its OSPB (Haladjian and Pylyshyn 2008), as would be expected if the tasks engage the same mechanism. There is, moreover, theoretical precedent for identifying the representations used in MOT with those used in object reviewing (Kahneman et al. 1992; Carey and Xu 2001; Pylyshyn 2007, 37–40). Accordingly, while multisystem accounts are available, they have questionable support.

Perhaps, however, the OF-system's processing principles differ from context to context. It might rely on the restrictive principles in some cases but not others. In this vein, Mitroff, Arita, and Fleck (2009) found that objects defined by subjective contours elicited OSPBs during blocks where they were the only objects presented but not when intermixed with trials involving physically defined objects. It is an empirical question whether such con-

6. Thanks to two anonymous reviewers for raising this possibility.

7. However, Burge (2010) explicitly claims that the strong restrictive view characterizes visual tracking. As such, MOT evidence alone challenges this position.

textual effects carry over to, say, collections and parts. If so, then a context-relativized account of the OF-system's principles may be viable.

If the OF-system picks out parts, collections, and shadows in addition to Spelke-objects, why call it an *object* file system at all? Granted, if "object" is interpreted to mean *Spelke-object*, this label may be misleading. However, the crucial question is whether the system responsible for tracking Spelke-objects is also responsible for tracking 2-D entities, parts, and collections. How to label this system is a further matter.

4. The Weak Restrictive View. I have considered whether the OF-system critically relies on attributing 3-D, cohesion, and boundedness. I have argued that the evidence cited for this view is inadequate and that further evidence raises problems for the view. This challenges the strong restrictive view, but does not directly target the weak restrictive view, which claims only that the OF-system has the proper function of selecting Spelke-objects. Might the weak restrictive view be true, even if the strong restrictive view is false?

Suppose that the strong restrictive view is false. The proponent of the weak restrictive view then needs to argue that the OF-system has the function of selecting Spelke-objects despite readily picking out non-Spelke-objects. One suggestion would be that object files are selectively tuned to Spelke-objects simply because the majority of entities they pick out in natural environments are Spelke-objects. However, this sort of proposal faces familiar problems (Fodor 1987, 99–104). For example, even if the majority of entities we pick out in natural environments are Spelke-objects, an even greater majority are members of a broader kind that includes Spelke-objects alongside certain 2-D entities (shadows or reflections), collections, and parts. So why think that object files are selectively tuned to Spelke-objects, rather than to some broader kind?

A more promising suggestion can be gleaned from Burge (2010). Burge holds that tracking relies on attributing the kind *body* and that a precondition for representing something as a body is that we represent it as 3-D, cohesive, and bounded. This entails the strong restrictive view. However, Burge also holds there are further properties constitutive of being a body that we need not attribute during tracking—namely, solidity (465–68). The reason, he argues, is that nonsolid entities that satisfy the remaining restrictive principles have not figured in biological explanations of our activities. Tracking can be keyed to solid entities even if it does not rely on attributing solidity, so long as only solid entities figure in biological explanations of our activities. Perhaps we should apply this teleological treatment of solidity to the 3-D, cohesion, and boundedness properties as well.

I believe the most compelling version of the weak restrictive view would combine this teleological approach with the claim that the OF-system tracks

objects by attributing perceptual organization cues. On this proposal, the restrictive principles characterize the kind of entity that the OF-system has the proper function of tracking, while perceptual organization principles characterize the properties that it typically attributes during tracking. This attempts to combine the insights of the restrictive and permissive views.

Nevertheless, assuming that we can readily track non-Spelke-objects without representing them as Spelke-objects, the weak restrictive view is dubious. Contrast the current situation with a classic case in which a system picks out entities it does not have the function of picking out. The frog's "bug detector" fires in response to any small, dark object moving through its field of vision. Nonetheless, many claim that the detector is tuned to flies, rather than pellets. Its firing causes tongue-flicking, which is only adaptive when directed at sources of nourishment. Thus, since flies plausibly figure in biological explanations of tongue-flicking, while pellets do not, the bug detector functions to detect flies.

Note that although tongue-flicking is fast and reflexive, this is not why the bug detector is plausibly tuned to flies. Rather, the reason is that those downstream behaviors caused by the bug detector are only successful given a fly. This might have obtained even if tongue-flicking were controlled by a slow, voluntary mechanism (although this would be far less efficient). The mechanism would perform its function in response to flies but not pellets.

Now consider cases of tracking non-Spelke-objects. We find two important disanalogies. First, there is an explanation of why the bug detector "misfires" in response to black pellets: The frog is unable to discriminate pellets from flies due to lack of visual acuity. But we plausibly can discriminate parts, collections, and shadows from Spelke-objects. So why should the OF-system "misfire" in response to these entities?

Second, we have a compelling story about why the bug detector only performs its function when it detects a fly—tongue-flicking is only successful given a fly. But no such story seems available for the tracking of collections, parts, or 2-D entities. Tracking such things is often adaptive. Tracking a swarm of bees can help us evade it. Selecting the handle of a mug can help us grasp it. And the ability to track the shadows of creatures flying overhead plausibly helped our evolutionary ancestors to avoid predation.⁸ These entities are both trackable and plausibly relevant in biological explanations of our activities.

Thus, if the strong restrictive view is false, then the weak restrictive view is probably false too. There are non-Spelke-objects that it is in our biological interest to select and track. If the OF-system tracks them, then it likely performs its function when doing so.

8. Cast shadows elicit alarm calls in squirrel monkeys (Newman 1985, 104).

I want to consider one more argument, due to Dickie (2010), for the claim that object files are selectively keyed to “ordinary” objects. Dickie proposes that every object file has a “Governing Conception,” characterized as a “cluster of dispositions that determines which rationally driven operations the subject will perform in maintaining the file” (224). Specifically, we are rationally disposed to maintain files through certain changes in the entities they pick out but not through others. Dickie proposes that an object file refers to an entity *O* only if its Governing Conception is matched to *O*’s category. Roughly, the file must be maintained only through changes that something of *O*’s category can undergo. Given this, she argues that object files can pick out ordinary objects, but not shadows or parts, because object files are systematically discarded in response to cues that an object is not ordinary.

Nevertheless, if by “ordinary object” Dickie means “Spelke-object,” then this account faces some of the same difficulties discussed earlier. There is evidence that perceivers can select and track objects despite cues inconsistent with cohesion and boundedness. Thus, it is questionable whether object files’ Governing Conceptions really are matched to Spelke-objects. Perhaps, however, Dickie is using “ordinary object” in a different way.⁹ If so, then her view is distinct from the positions addressed here and should be evaluated separately.

5. Conclusion. A popular view, which I have called the *restrictive view*, claims that object files are selectively keyed to 3-D, cohesive, bounded entities (Spelke-objects). I have argued that the evidential support for this view is unpersuasive.

I first considered the *strong restrictive view*, according to which the OF-system relies on attributing 3-D, cohesion, and boundedness during tracking. I argued that the evidence garnered in support of this view can be explained equally well by a *permissive view*, according to which object files rely on well-known cues to perceptual organization. The evidence for cohesion attribution can be explained by the attribution of proximity and common fate. The evidence for boundedness attribution shows only that we cannot track arbitrary parts of objects, not that we cannot track parsable parts. Work on the selection and tracking of shadows, groups, and parts provides further evidence against the strong restrictive view and in support of the permissive view.

Next, I considered the *weak restrictive view*, according to which object files have the proper function of picking out Spelke-objects. I argued that if the strong restrictive view is false, then the weak restrictive view is likely

9. Dickie notes that ordinary objects are spatiotemporally continuous and causally unified and that their parts stand in stable relations to each other. However, she does not explicitly appeal to the cohesion or boundedness criteria.

false too. There are many non-Spelke-objects that it is adaptive to track. If the OF-system tracks them, then it plausibly performs its function when doing so.

This raises some further questions. Adults conceptually distinguish among bodies, parts, and collections. Children believe that certain objects can persist despite changes in the substances from which they are constituted (Rips and Hespos 2015), and they generally prefer to apply novel words to “whole” objects, rather than to parts (Markman 1990). If object files are not strictly governed by the body/nonbody dichotomy, where do these distinctions come from?

Note that even if the permissive view is correct, this does not entail that Spelke-objects play no role in perception. The distinction between Spelke-objects and non-Spelke-objects may be perceptually salient even if it does not mark the boundaries of what object files track. Indeed, cohesion and boundedness are topological properties, and there is strong evidence that vision represents topology (Chen 2005; Green 2017). Thus, the property of being a Spelke-object may still be visually represented and made available for judgment and language acquisition.

If object files are not selectively keyed to Spelke-objects, then what are they keyed to? One possibility is that there is no easily specifiable category to which they are keyed. Another possibility is that they are keyed to a broader category than Spelke-objects. Note that Spelke-objects, perceptual groups, and parsable parts all tend to exhibit a high degree of internal dependence. Their parts have stronger probabilistic dependencies on one another than on the parts of other entities. This is clear in the case of motion. Parts of my arm are better motion correlated with one another than with parts of my torso. Likewise, a goose is better motion correlated with another from the same flock than one from a different flock. One possibility, then, is that it is adaptive for us to pick out things that exhibit such internal dependence, and the OF-system exploits the rich assortment of cues to internal dependence that natural scenes offer.

REFERENCES

- Bahrami, Bahador. 2003. “Object Property Encoding and Change Blindness in Multiple Object Tracking.” *Visual Cognition* 10:949–63.
- Barenholtz, Elan, and Jacob Feldman. 2003. “Visual Comparisons within and between Object Parts: Evidence for a Single-Part Superiority Effect.” *Vision Research* 43:1655–66.
- Bertamini, Marco. 2006. “Who Owns the Contour of a Visual Hole?” *Perception* 35:883–94.
- Bhatt, Ramesh S., Angela Hayden, Ashley Kangas, Nicole Zieber, and Jane E. Joseph. 2010. “Part Perception in Infancy: Sensitivity to the Short-Cut Rule.” *Attention, Perception, and Psychophysics* 72:1070–78.
- Brady, Timothy F., and Joshua B. Tenenbaum. 2013. “A Probabilistic Model of Visual Working Memory: Incorporating Higher Order Regularities into Working Memory Capacity Estimates.” *Psychological Review* 120:85–109.

- Brovold, Amanda, and Rick Grush. 2012. "Towards an (Improved) Interdisciplinary Investigation of Demonstrative Reference." In *Perception, Realism, and the Problem of Reference*, ed. Athanasios Raftopoulos and Peter Machamer, 11–42. Cambridge: Cambridge University Press.
- Burge, Tyler. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.
- Cacchione, Trix. 2013. "The Foundations of Object Permanence: Does Perceived Cohesion Determine Infants' Appreciation of the Continuous Existence of Material Objects?" *Cognition* 128:397–406.
- Cameron, Patricia A., and Gordon G. Gallup. 1988. "Shadow Recognition in Human Infants." *Infant Behavior and Development* 11:465–71.
- Carey, Susan. 2009. *The Origin of Concepts*. Oxford: Oxford University Press.
- Carey, Susan, and Fei Xu. 2001. "Infant Knowledge of Objects: Beyond Object Files and Object Tracking." *Cognition* 80:179–213.
- Chen, Lin. 2005. "The Topological Approach to Perceptual Organization." *Visual Cognition* 12:553–637.
- Chen, Zhe. 2012. "Object-Based Attention: A Tutorial Review." *Attention, Perception, and Psychophysics* 74:784–802.
- Cherries, Erik W., Stephen R. Mitroff, Karen Wynn, and Brian J. Scholl. 2008. "Cohesion as a Constraint on Object Persistence in Infancy." *Developmental Science* 11:427–32.
- de-Wit, Lee, David Milner, and Robert Kentridge. 2012. "Shadows Remain Segmented as Selectable Regions in Object-Based Attention Paradigms." *i-Perception* 3:150–58.
- Dickie, Imogen. 2010. "We Are Acquainted with Ordinary Things." In *New Essays on Singular Thought*, ed. Robin Jeshion, 213–46. Oxford: Oxford University Press.
- Drew, Trafton, Todd S. Horowitz, Jeremy M. Wolfe, and Edward K. Vogel. 2011. "Delineating the Neural Signatures of Tracking Spatial Position and Working Memory during Attentive Tracking." *Journal of Neuroscience* 31:659–68.
- Feldman, Jacob. 2007. "Formation of Visual 'Objects' in the Early Computation of Spatial Relations." *Perception and Psychophysics* 69:816–27.
- Fodor, Jerry A. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, Jerry A., and Zenon W. Pylyshyn. 2015. *Minds without Meanings: An Essay on the Content of Concepts*. Cambridge, MA: MIT Press.
- Fougnie, Daryl, and Rene Marois. 2009. "Attentive Tracking Disrupts Feature Binding in Visual Working Memory." *Visual Cognition* 17:48–66.
- Giralt, Nuria, and Paul Bloom. 2000. "How Special Are Objects? Children's Reasoning about Objects, Parts, and Holes." *Psychological Science* 11:497–501.
- Green, E. J. 2017. "A Layered View of Shape Perception." *British Journal for the Philosophy of Science* 68 (2): 355–87. doi:10.1093/bjps/axv042.
- Haladjian, Harry H., and Zenon W. Pylyshyn. 2008. "Object-Specific Preview Benefit Enhanced during Explicit Multiple Object Tracking." *Journal of Vision* 8:497.
- Hoffman, Donald D., and Whitman A. Richards. 1984. "Parts of Recognition." *Cognition* 18:65–96.
- Hollingworth, Andrew, and Ian P. Rasmussen. 2010. "Binding Objects to Locations: The Relationship between Object Files and Visual Working Memory." *Journal of Experimental Psychology: Human Perception and Performance* 36:543–64.
- Horowitz, Todd S., and Yoana Kuzmova. 2011. "Can We Track Holes?" *Vision Research* 51:1013–21.
- Howe, Piers D., Natalie C. Incledon, and Daniel R. Little. 2012. "Can Attention Be Confined to Just Part of a Moving Object? Revisiting Target-Distractor Merging in Multiple Object Tracking." *PLOS One* 7:e41491.
- Howe, Piers D. L., Alex O. Holcombe, Mark D. Lapierre, and Simon J. Cropper. 2013. "Visually Tracking and Localizing Expanding and Contracting Objects." *Perception* 42:1281–1300.
- Huntley-Fenner, Gavin, Susan Carey, and Andrea Solimando. 2002. "Objects Are Individuals but Stuff Doesn't Count: Perceived Rigidity and Cohesiveness Influence Infants' Representations of Small Groups of Entities." *Cognition* 85:203–21.
- Kahneman, Daniel, Anne Treisman, and Brian J. Gibbs. 1992. "The Reviewing of Object Files: Object-Specific Integration of Information." *Cognitive Psychology* 24:175–219.
- Kellman, Philip J., and Elizabeth S. Spelke. 1983. "Perception of Partly Occluded Objects in Infancy." *Cognitive Psychology* 15:483–524.

- Kellman, Philip J., Elizabeth S. Spelke, and Kenneth R. Short. 1986. "Infant Perception of Object Unity from Translatory Motion in Depth and Vertical Translation." *Child Development* 57:72–86.
- Kimchi, Ruth. 2009. "Perceptual Organization and Visual Attention." *Progress in Brain Research* 176:15–33.
- Luck, Steven J., and Edward K. Vogel. 1997. "The Capacity of Visual Working Memory for Features and Conjunctions." *Nature* 309:279–81.
- Markman, Ellen M. 1990. "Constraints Children Place on Word Meanings." *Cognitive Science* 14:57–77.
- Mitroff, Stephen R., Jason T. Arita, and Mathias S. Fleck. 2009. "Staying in Bounds: Contextual Constraints on Object-File Coherence." *Visual Cognition* 17:195–211.
- Mitroff, Stephen R., Brian J. Scholl, and Karen Wynn. 2004. "Divide and Conquer: How Object Files Adapt When a Persisting Object Splits into Two." *Psychological Science* 15:420–25.
- . 2005. "One Plus One Equals One: The Effects of Merging on Object Files." Poster presented at the annual meeting of the Psychonomic Society, Toronto.
- Murez, Michael, and François Recanatani. 2016. "Mental Files: An Introduction." *Review of Philosophy and Psychology* 7:265–81.
- Newman, John D. 1985. "Squirrel Monkey Communication." In *Handbook of Squirrel Monkey Research*, ed. Leonard A. Rosenblum and Christopher L. Coe, 99–126. New York: Springer.
- Olson, Ingrid R., and Yuhong Jiang. 2002. "Is Visual Short-Term Memory Object Based? Rejection of the 'Strong-Object' Hypothesis." *Perception and Psychophysics* 64:1055–67.
- Palmer, Stephen E., Janet Davis, Rolf Nelson, and Irvin Rock. 2008. "Figure-Ground Effects on Shape Memory for Objects versus Holes." *Perception* 37:1569–86.
- Palmer, Stephen E., and Irvin Rock. 1994. "Rethinking Perceptual Organization: The Role of Uniform Connectedness." *Psychonomic Bulletin and Review* 1:29–55.
- Peterson, Dwight J., and Marian E. Berryhill. 2013. "The Gestalt Principle of Similarity Benefits Visual Working Memory." *Psychonomic Bulletin and Review* 20:1282–89.
- Porter, Katherine B., Veronica Mazza, Annie Garofalo, and Alfonso Caramazza. 2016. "Visual Object Individuation Occurs over Object Wholes, Parts, and Even Holes." *Attention, Perception, and Psychophysics* 78:1145–62.
- Pylyshyn, Zenon W. 2007. *Things and Places: How the Mind Connects with the World*. Cambridge, MA: MIT Press.
- Pylyshyn, Zenon W., and Ron W. Storm. 1988. "Tracking Multiple Independent Targets: Evidence for a Parallel Tracking Mechanism." *Spatial Vision* 3:179–97.
- Quinn, Paul C., and Ramesh S. Bhatt. 2005. "Learning Perceptual Organization in Infancy." *Psychological Science* 16:511–15.
- Rehman, Anis Ur, Ken Kihara, Akiko Matsumoto, and Sakuichi Ohtsuka. 2015. "Attentive Tracking of Moving Objects in Real 3D Space." *Vision Research* 109:1–10.
- Rips, Lance J., and Susan J. Hespos. 2015. "Divisions of the Physical World: Concepts of Objects and Substances." *Psychological Bulletin* 141:786–811.
- Scholl, Brian J., and Alan M. Leslie. 1999. "Explaining the Infant's Object Concept: Beyond the Perception/Cognition Dichotomy." In *What Is Cognitive Science?* ed. Ernest Lepore and Zenon W. Pylyshyn, 26–73. Oxford: Blackwell.
- Scholl, Brian J., and Zenon W. Pylyshyn. 1999. "Tracking Multiple Items through Occlusion: Clues to Visual Objecthood." *Cognitive Psychology* 38:259–90.
- Scholl, Brian J., Zenon W. Pylyshyn, and Jacob Feldman. 2001. "What Is a Visual Object? Evidence from Target Merging in Multi-Element Tracking." *Cognition* 80:159–77.
- Singh, Manish, and Donald D. Hoffman. 2001. "Part-Based Representations of Visual Shape and Implications for Visual Cognition." *Advances in Psychology* 130:401–59.
- Singh, Manish, Gregory D. Seyranian, and Donald D. Hoffman. 1999. "Parsing Silhouettes: The Short-Cut Rule." *Perception and Psychophysics* 61:636–60.
- Spelke, Elizabeth S. 1990. "Principles of Object Perception." *Cognitive Science* 14:29–56.
- . 1994. "Initial Knowledge: Six Suggestions." *Cognition* 50:443–47.
- Spelke, Elizabeth S., Karen Breinlinger, Janet Macomber, and Kristen Jacobson. 1992. "Origins of Knowledge." *Psychological Review* 99:605–32.
- St. Clair, Rebecca, Markus Huff, and Adriane E. Seiffert. 2010. "Conflicting Motion Information Impairs Multiple Object Tracking." *Journal of Vision* 10:1–13.

- Tversky, Barbara. 1989. "Parts, Partonomies, and Taxonomies." *Developmental Psychology* 25:983–95.
- vanMarle, Kristy, and Brian J. Scholl. 2003. "Attentive Tracking of Objects versus Substances." *Psychological Science* 14:498–504.
- Wagemans, Johan, James H. Elder, Michael Kubovy, Stephen E. Palmer, Mary A. Peterson, Manish Singh, and Rüdiger von der Heydt. 2012. "A Century of Gestalt Psychology in Visual Perception." Pt. 1, "Perceptual Grouping and Figure-Ground Organization." *Psychological Bulletin* 138:1172–1217.
- Webb, Ben S., Timothy Ledgeway, and Francesca Rocchi. 2011. "Neural Computations Governing Spatiotemporal Pooling of Visual Motion Signals in Humans." *Journal of Neuroscience* 31:4917–25.
- Woodman, Geoffrey F., Shaun P. Vecera, and Steven J. Luck. 2003. "Perceptual Organization Influences Visual Working Memory." *Psychonomic Bulletin and Review* 10:80–87.
- Wynn, Karen, Paul Bloom, and Wen-Chi Chiang. 2002. "Enumeration of Collective Entities by 5-Month-Old Infants." *Cognition* 83: B55–B62.
- Xu, Yaoda. 2002. "Encoding Color and Shape from Different Parts of an Object in Visual Short-Term Memory." *Perception and Psychophysics* 64:1260–80.
- Zhou, Ke, Huan Luo, Tiangang Zhou, Yan Zhuo, and Lin Chen. 2010. "Topological Change Disturbs Object Continuity in Attentive Tracking." *Proceedings of the National Academy of Sciences* 107:21920–24.
- Zosh, Jennifer M., and Lisa Feigenson. 2009. "Beyond 'What' and 'How Many': Capacity, Complexity, and Resolution of Infants' Object Representations." In *The Origins of Object Knowledge*, ed. Bruce M. Hood and Laurie R. Santos, 25–51. Oxford: Oxford University Press.