Wynn, K., Bloom, P., & Chiang, W-C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition*, *83*, B55–B62.

Xu, F. (2003). Numerosity discrimination in infants: evidence for two systems of representations. *Cognition*, *89*, B15–B25.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, *30*, 111–153.

Xu, F., Carey, S., & Welch, J. (1999). Infants' ability to use object kind information for object individuation. *Cognition*, *70*, 137–166.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11.

Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: a survey. *ACM Computing Surveys*, *38*, 13.

Chapter 6

Spatiotemporal priority as a fundamental principle of object persistence

Jonathan I. Flombaum, Brian J. Scholl, & Laurie R. Santos

6.1 Introduction

The impoverished and rapidly changing stimulation on the retina looks very different from the stable world of discrete persisting objects that populate our visual experience. To get from the features on the retina to the objects that we experience, the visual system must solve several correspondence problems. One of these problems has to do with *sameness*: the visual system must decide whether each bit of stimulation reflects an object that has already been encountered (which might occasion the *updating* of an existing object representation) or a new one (which might occasion the *creation* of a new object representation). This problem of object persistence has been studied with a wide array of visual phenomena and paradigms, and in several disciplines in cognitive science—including vision science, developmental psychology, and comparative cognition.

The study of object persistence in these different fields has progressed largely independently. Yet strikingly, they have converged on a core principle that guides the creation and maintenance of persisting object representations: the principle of *spatiotemporal priority*. When identifying objects as the same individuals over the time, the visual system appears to rely on their spatiotemporal histories—that is, where, when, and how they were encountered—to a greater degree than their visual surface features. In this chapter, we review the many contexts in which spatiotemporal priority drives computations of object persistence, and we propose explanations at several levels for why spatiotemporal priority plays this dominant role.

6.1.1 The problem of persistence

Suppose that you have just spent several hours at a large suburban shopping mall, and you must now retrieve your Prius from the parking lot. When you exit the mall, a sea of cars overwhelms you. How will you decide which is yours? This is a type of correspondence problem: you must find the car that corresponds to the one in which you arrived. These kinds of problems are extremely prevalent in moment-to-moment visual experience, but because they are solved so efficiently, we may not always realize it. Indeed, the problem of object persistence is fundamentally a problem of correspondence: our visual system must constantly decide whether current stimulation reflects a novel object, or whether it corresponds to an object that was already encountered a moment ago (Scholl & Flombaum, in press). If we couldn't solve this type of problem, visual experience would be incoherent. This is a problem in part due to the sheer information load involved, and in part because a previously experienced object may be reencountered—even a moment later—in a different location (perhaps after a visual interruption such as occlusion), and/or with different visual surface features. How are these challenges overcome? How do we readily see objects as the *same* enduring individuals from moment to moment?

In the case of finding your Prius in the mall's crowded parking lot, there are at least two ways to solve the problem: you might scan the lot for a car that *looks* like yours or you might try to remember where exactly you parked it. These same two solutions are available to meet the challenges of persistence in online perception. First, the visual system can rely on what objects look like to reidentify them: 'this red square must be the same red square that I just saw a moment ago.' Second, the visual system can rely on where and when various objects appear: 'this object *here* must be same object that was previously over *there* but was moving fast in exactly this direction.' Research across a wide array of disciplines and phenomena converges on a simple pattern of results: the second kind of solution trumps the first. That is, the correspondences favored by spatiotemporal considerations will almost always trump those favored by featural similarity. We will term this the *principle of spatiotemporal priority* (see also Scholl, 2001b).

This bias may seem intuitively surprising. If in the mall parking lot you arrived at the location where you left your Prius but found an Escalade in that location, you would not naturally assume that your car had somehow changed its appearance while you were busy shopping. Yet, the visual system seems comfortable making analogous assumptions. In fact, because of spatiotemporal priority, we readily see impossible featural transformations in a variety of visual phenomena, and infants and nonhuman primates often make rather

large errors when keeping track of objects because they discount objects' surface properties. In the next two sections, we review this evidence.

6.2 Spatiotemporal priority in adult visual cognition

6.2.1 Amodal completion

Among the most extensively studied correspondence problems in visual perception is the phenomenon of amodal completion—our ability to perceive objects as completing behind occluders (Kanizsa, 1979; Michotte et al., 1964/1991; Wertheimer, 1912/1961). For example, we naturally see two of the line segments in Fig. 6.1a as two ends of the same line, occluded by the box. And, despite the 'ink spills', we can naturally bind together the many segments in Fig. 6.1b in order to recognize the necker cube. This is a spatial effect, of course, rather than a spatiotemporal effect, but it illustrates the character of the problem and its solutions. In particular, note that the completion in Fig. 6.1a cannot be based only on the surface similarity of the line segments, because there are many other identical segments that are not bound to one another. Instead, visual completion in such spatial contexts is driven by subtle factors of how the line segments are aligned relative to each other (e.g., Kellman & Shipley, 1991; Singh & Fulvio, 2005). One way to explain such effects is in terms of the visual system's tendency to avoid coincidences: in these cases, the occluded fragments are aligned in a way that would be highly coincidental if they were in fact unrelated. This is a conclusion that the visual system is unwilling to accept, and so it settles on an interpretation of occluded objects, which 'explains away' the coincidence.

This emphasis on spatial alignment rather than featural similarity is notable in two ways. First, it means that segments may be bound into the same (occluded) object representations even when those segments look different (as in Fig. 6.1c; e.g., Walker & Davies, 2003); inconsistencies in surface appearances, in other words, do not preclude completion. Second, an emphasis on spatial alignment means that we almost always perceive the most direct link between unoccluded segments, as in Fig. 6.1d and 6.1e (adapted from examples in Kanizsa, 1982). These cases also highlight the nature of such processing as fundamentally perceptual rather than cognitive—because it might be more natural to *think* that regular patterns in these images simply continue behind the occluders, although this clearly not what you see.

6.2.2 Amodal integration and the tunnel effect

The previous section lingered on a type of *spatial* correspondence problem largely because the nature of the problem—and its solution—is entirely

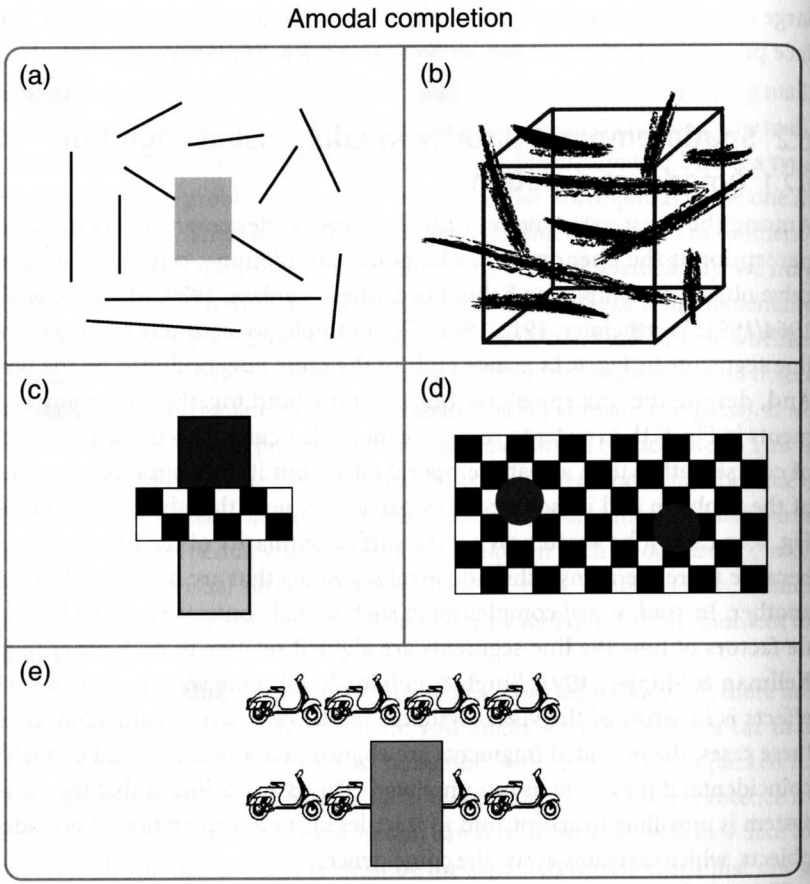

Fig. 6.1 Examples of amodal completion. (a) We readily bind segments together to perceive whole objects despite occlusion. (b) This phenomenon is robust, and we can even recognize large objects that are occluded in multiple places. (c) Completion can occur on the basis of spatial alignment, despite differences in surface features. (d) The binding process often ignores featural regularities, for example, leading us to see the circles as occluding black and white crosses, even though it makes more 'sense', cognitively, to interpret the entire background as a checkerboard (adapted from Kanizsa, 1982). (e) Similarly, we see a single 'anomalous' extended scooter behind the occluder, even though it makes more 'sense', cognitively, to interpret all the scooters as identical (adapted from Kanizsa, 1982).

analogous in spatio*temporal* contexts that require correspondences among objects encountered at different times and places. Inspired by the phenomenon of amodal completion, Luke Burke and Albert Michotte realized in the 1950s that we do not only see *parts* of a static object that are currently occluded but we also readily see *entire* objects as persisting through periods of occlusion. Here is their elegant description of the problem:

> Let us suppose a cyclist has to pass behind some stationary object in the centre of the street in which we are standing. He moves towards the object, disappears from view, and a moment later reappears at the other side. In this every day fact we have … an intriguing problem in the psychology of perception. Every observer is convinced that cyclist has been continuously in motion. Is this merely a matter of belief or of knowledge based on past experience or is the continuity of the movement actually 'seen' by the observer?
>
> (Burke, 1959, p. 122)

To answer these questions, they constructed demonstrations of various objects passing at different speeds behind occluders of various lengths, and they simply asked observers to report their percepts. The responses they collected (and indeed, their own visual experiences) confirmed that we really do see such events in terms of the (occluded) motion of a single persisting object. This type of dynamic 'amodal integration' is the spatiotemporal analogue of spatial amodal completion. And similar to amodal completion, the phenomenon is exquisitely sensitive to the details: when the conditions are not quite right for amodal integration, observers will instead see two distinct objects—one which disappears behind an occluder and another that later emerges. Thus, the computations that underlie amodal integration influence the contents of our mental experience—determining whether we perceive one object or two (Burke, 1952; Michotte et al., 1964/1991).

Burke and Michotte also made two other key observations about amodal integration. First, they observed that we may often see a single occluded moving object even if the object that enters an occluder looks very different from the one that later emerges. For example, the occluding object might be light and the disoccluding object might be dark—but we would (often irresistibly) perceive this in terms of a single object that *changed* from light to dark, rather than as two separate objects (Fig. 6.2a). They termed this phenomenon of amodal integration despite featural heterogeneity the *tunnel effect*. Conversely, they noted that we *fail* to perceive an amodally integrated object—even if its features do *not* change—if the disoccluding object emerges at the wrong time (Fig. 6.2b) or place (Fig. 6.2c) given its initial trajectory. To summarize both observations, spatiotemporal factors such as continuity trump objects' surface features during computations of object persistence.

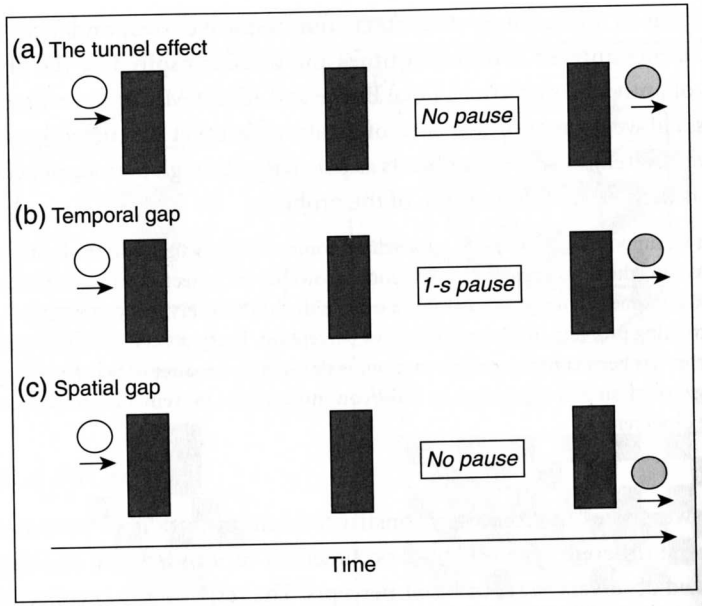

Fig. 6.2 The tunnel effect. (a) We readily perceive a persisting object through occlusion, even if the pre- and post-occlusion views have distinct features (e.g., color), provided that the disoccluding object appears at the right time and place. The percept does not obtain, however, when the trajectories are discontinuous as a result of either a temporal gap (b) or spatial gap (c).

This spatiotemporal bias is more than just a quirk in perception. Instead, it appears to be a central feature of our underlying representations of persisting objects—and as a result, spatiotemporal priority can impact the downstream processes and behaviors that rely on these representations. For example, in a study of the influence of object persistence on visual memory, we recently discovered that an amodally integrated percept facilitates change detection of objects' colors (Flombaum & Scholl, 2006). Our displays included several independent dynamic events in which an object disappeared behind an occluder, and then an object emerged from the other side of the occluder, as depicted in Fig. 6.3. (These effects are inherently dynamic, and readers can view example displays at http://www.yale.edu/perception/tunnel/.)

Observers' task was simply to press a key when they detected a color change (i.e., when a disoccluding object was a different color from an occluding object) in any of the simultaneously moving events. Subjects were more accurate in these judgments under conditions when the tunnel effect obtained—that is, when the two objects were bound into the same underlying object representation—compared with when two objects were perceived (such as

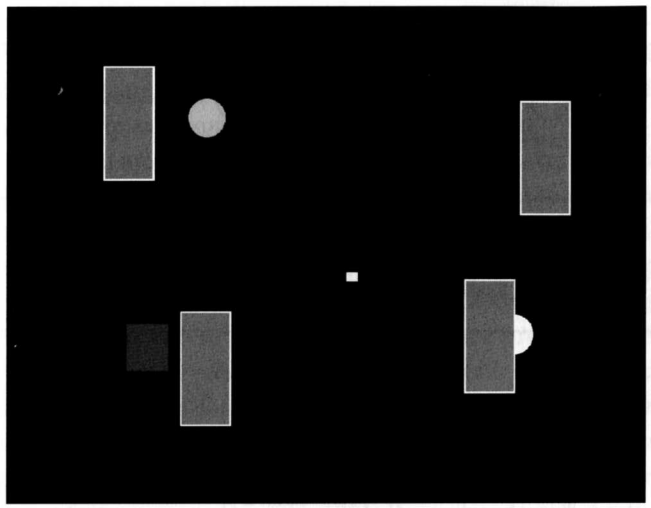

Fig. 6.3 The tunnel effect and visual memory. A static snapshot (not to scale) of a sample display in Experiment 1 from Flombaum and Scholl (2006). Readers can view dynamic animations of the conditions from each experiment online at http://www.yale.edu/perception/tunnel/.

cases involving spatial and temporal gaps, as in Fig. 6.2b and 6.2c). In other words, visual memory was better when the color change involved *updating* a single representation, rather than comparing two distinct representations. These studies indicate that spatiotemporal priority determines not just what we see, but also how well we remember it. Spatiotemporal priority is evidently not just a heuristic for perceiving persisting objects, but it underlies the way that we represent objects from moment to moment.

6.2.3 Apparent motion

Perhaps the most extensively studied example of spatiotemporal priority in visual cognition is apparent motion. In apparent motion, we readily see a continuous motion trajectory when an object appears in one location, disappears, and then an object quickly appears in another nearby location (Anstis, 1980; Wertheimer, 1912/1961; see Fig. 6.4a). This is another example of a type of coincidence avoidance in perception: the two spatiotemporally proximate flashes are judged unlikely to be independent, and the result is a percept that binds them together.

As in the tunnel effect and amodal completion, spatiotemporal factors primarily determine whether and when this percept obtains. We can readily see apparent motion when the two discrete flashes involve objects that

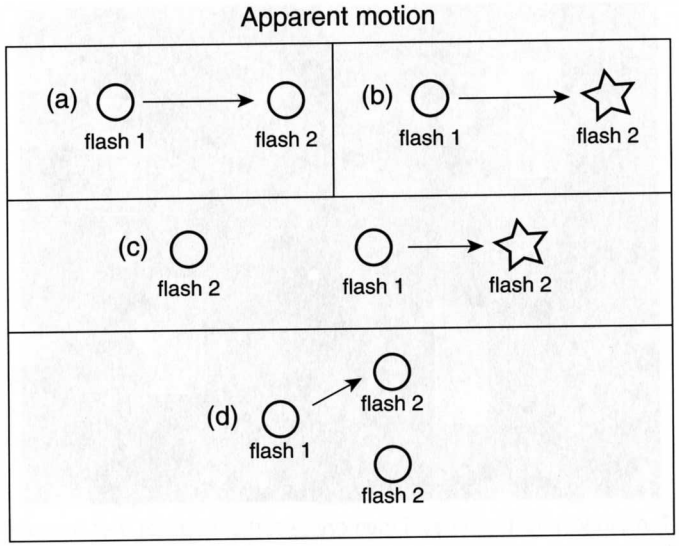

Fig. 6.4 Apparent motion. (a) In apparent motion, we are presented with two discrete stationary flashes, but we see a single object move between the flashed locations. (b) This percept obtains even if the two flashes look dissimilar, and (c) distance trumps featural similarity when matches favoring these factors are both available. (d) Apparent motion also obeys the spatiotemporal rule of solidity, or object integrity: when two equidistant correspondences are available, we see only a single motion trajectory.

look very different from each other, so long as the flashes are close enough in time and space (Kolers, 1972; Fig. 6.4b). Conversely, we will fail to see apparent motion when the flashes are not sufficiently spatiotemporally proximate, even if they look identical (e.g., Dawson, 1991). Moreover, distance trumps surface appearance. When an object disappears and then two objects appear simultaneously, another correspondence problem arises: which of the two new objects (if either) is the same individual as the one that disappeared? The visual system's solution to this correspondence problem depends on distance: we will see apparent motion linking the disappeared object with the one that appears closer to the disappearance, even if the more distant one would avoid a large featural transformation, as in Fig. 6.4c (Kolers & Pomerantz, 1971). In some experiments, there appears to be no preference at all for featurally consistent versus inconsistent percepts, even with all other spatiotemporal factors held equal (Burt & Sperling, 1981). Thus, at best, surface features play a weak, secondary, and indirect role in generating apparent motion percepts. Changes in surface features are not sufficient to produce motion on their own; they must be accompanied by spatiotemporal changes (Navon, 1976). When they

do have an impact, they do so merely by modulating sensitivity to distance (Shechter et al., 1988), and models of apparent motion have estimated the impact of surface features on motion correspondence is as much as 80 times weaker than the impact of spatiotemporal factors (Dawson, 1991).

In fact, successful models of apparent motion are best characterized in terms of a handful of simple spatiotemporal rules that also seem to apply in related phenomena such as amodal completion and integration (Dawson, 1991). In particular, because of the *nearest neighbor principle*, we tend to perceive apparent motion trajectories that minimize the distance traveled (Fig. 6.4c). Similarly, in displays where several correspondence problems need to be solved simultaneously, the *relative velocity principle* leads to a minimization of the amount of relative motion between corresponding pairs. In other words, global spatiotemporal properties in a display, not just local ones, also drive apparent motion—but once again, these computations function irrespective of surface features (e.g., Dawson, 1991; Ramachandran & Anstis, 1986). Finally, according to the *element integrity principle*, perception avoids generating correspondences whereby an element is split into parts (i.e., one element moves to two new locations simultaneously; see Fig. 6.4d), or whereby two elements fuse together (e.g., Dawson, 1991; Ullman, 1979). As we discuss later, a similar spatiotemporal principle, often called *cohesion*, seems to drive infants' reasoning about persisting objects.

As in amodal integration and the tunnel effect, the principle of spatiotemporal priority in apparent motion affects more than just what we see. In one relevant study, for example, observers had to detect repeated letters that appeared in two rapid serial visual presentation (RSVP) streams (Chun & Cavanagh, 1997). These two RSVP streams were not spatially stationary, however; instead, they constituted locations in two different apparent motion streams that passed by one another. Thus, as depicted in Fig. 6.5, letter repetitions could occur in a single apparent motion stream (making those letters the changing features of a single object over time)—or in two different apparent motion streams (making each cross-stream appearance of a repeated letter a feature of a different object). A common phenomenon in RSVP displays is that closely spaced repetitions are difficult to detect and perceive—a phenomenon known as *repetition blindness* (Kanwisher, 1987). In this experiment, however, observers experienced more repetition blindness for repeated letters that appeared in a single apparent motion stream, compared with those that appeared in distinct motion streams—even though the timing and spacing of these repetitions were identical in all conditions (Chun & Cavanagh, 1997). Thus, the spatiotemporal rules that guide the experience of apparent motion can also impact how we attend to the world

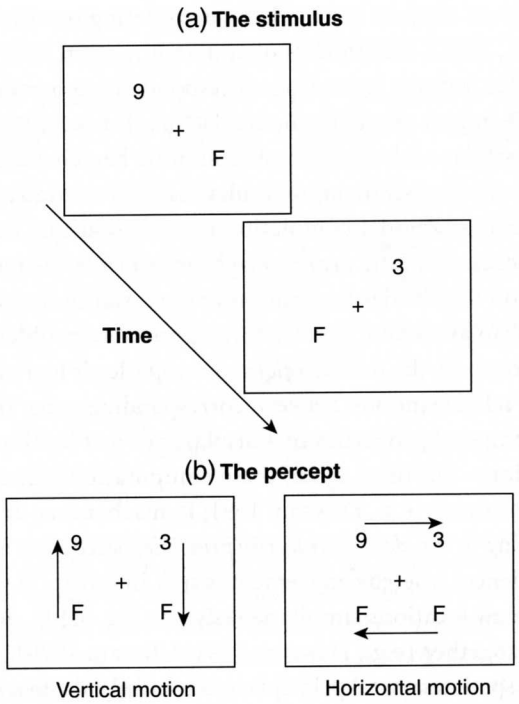

Fig. 6.5 Apparent motion and repetition blindness. (a) Chun and Cavanagh (1997) presented observers with motion quartets that could be biased toward a percept of either vertical or horizontal motion by the preceding apparent motion streams. (b) These observers detected fewer letter repetitions (i.e., they exhibited more repetition blindness) when repeated letters appeared in the same apparent motion stream (e.g., as in the 'Horizontal motion' panel), compared with when they appeared in different apparent motion streams (e.g., as in the 'Vertical motion' panel). (Adapted from Chun & Cavanagh, 1997)

and can ultimately even influence whether we will be visually aware of an object in the first place.

6.2.4 Illusory conjunctions

Similar to many of the challenges faced by the visual system, the challenge of object persistence is at root a type of inverse problem. Discrete and persisting objects exist in the real world, but this information is lost during optical transmission and must be reconstructed. One idiosyncratic aspect of this challenge is known as the *binding problem*. Objects in the real world possess a variety of features such as color, shape, size, and position. During the earliest phases of visual processing, these features are processed by distinct neural circuitry, and so this information must be later rebound in order to produce the impression

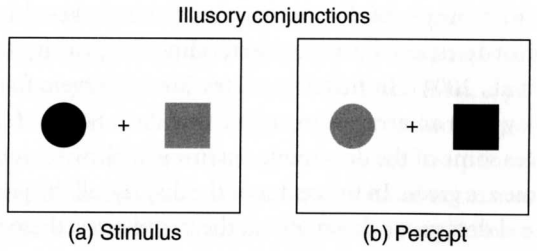

Fig. 6.6 Illusory conjunctions. Treisman and Schmidt (1982) demonstrated that under high attentional loads, observers sometimes misbind objects' surface features, such as lightness or color.

of coherent persisting individuals, with the right features attached to the right objects. In everyday life, it is difficult to recognize this challenge, because it is almost always met so successfully. Sometimes the binding operation does not work as accurately as it should, however, and we perceive *illusory conjunctions* between features and objects (Treisman & Schmidt, 1982; see Fig. 6.6).

An interesting aspect of such illusory conjunctions is that they appear to preserve spatiotemporal features at the expense of surface features. That is, the surface features, not the spatiotemporal features, seem capable of being bound to the wrong object. For example, if you are quickly presented with a red triangle on the left and a blue circle on the right, you might see a red circle on the left (misbinding shape to the wrong object), but you will not see a red triangle on the right (misbinding location to the wrong object). If all features were created equal—shape, color, and location, in this example—then these two illusory conjunctions would be equally likely. But all features are not created equal, and in the case of such illusory conjunctions, location seems privileged.

This same pattern occurs in more extreme cases of illusory conjunctions that arise in certain neuropsychological syndromes. For example, patients with Balint syndrome are frequently able to perceive only a single object at a time (for a review, see Rafal, 1997). Despite seeing only one object in a display filled with multiple objects, however, the single perceived object will occasionally serially adopt the surface properties of the other objects in the display—so that, for example, they might see a red circle that shifts to blue, then green, and so on (Humphreys et al., 2000; Robertson et al., 1997). Again, this is a case of a surface feature (color) floating free of an object. In contrast, spatial location does not seem to do this: these patients never report seeing a red circle in one place, and then having it shift to another location. It has even been suggested that Balint syndrome actually reflects an inability to represent more than a single spatial location at a time, which in turn results in the inability to perceive more than one object at a time (Rafal, 1997).

A more recent example of illusory conjunctions makes this type of spatiotemporal priority especially salient, extending the priority to direction of motion (Wu et al., 2004). In this demonstration, observers fixate the center of a large display with an array of many moving dots. Some of the dots move upward, whereas some of the dots move downward; likewise, some of the dots are red and some are green. In the center of the display, all the green dots move up and all the red dots move down. But at the peripheries, these motion directions are reversed: red dots move up and green dots move down. Nonetheless, observers perceive the peripheral dots as obeying the regularity set by the dots near fixation: they perceive *all* of the green dots in the display as moving up, and *all* the red dots as moving down.

This is a striking example of illusory conjunctions persisting over time and on a large scale, and it occurs because of a type of spatiotemporal priority: global motion coherence appears to be more important to the visual system than correctly locating green things and red things. Illusory conjunctions thus suggest that position appears privileged relative to color and other surface properties, and that it seems to form the scaffolding over which object representations are built. In other words, at least some surface features merely ride on the backs of object representations that are set up on the basis of spatiotemporal considerations. (This is a conclusion that we will return to in a discussion of 'object files'.)

6.3 Ontogenetic and phylogenetic origins of spatiotemporal priority

In the previous section, we reviewed evidence for spatiotemporal priority in adult visual cognition. In this section, we turn to the origins of this phenomenon. Does spatiotemporal priority develop only after extended experience with the behavior of visual objects, or does this bias emerge early in development? Similarly, is this phenomenon uniquely human, or is it shared more widely in the animal kingdom? Put differently, do infants and nonhuman animals experience objects as persisting individuals over time—and if so, do they prioritize spatiotemporal features when determining persistence?

6.3.1 Spatiotemporal priority in infant object cognition

The question of whether infants recognize that objects persist over time (even when they are not always in view) has played a prominent role in research on infant object cognition, ever since Piaget's time. Piaget famously claimed on the basis of observing infants' overt behaviors that they do not see objects as persisting through interruptions such as occlusion—that they lack what

he called *object permanence* (Piaget, 1954). It took nearly 30 years for clever researchers armed with new tools such as the measurement of looking times to prove him wrong (e.g., Baillargeon, 1986). Since then, a growing body of work suggests that infants recognize that objects continue to persist over time, and that their persisting object representations survive periods of occlusion and several sorts of manipulations during periods of occlusion (e.g., Cheries et al., 2006; Spelke et al., 1992; Wynn, 1992; Xu & Carey, 1996; for reviews, see Baillargeon, 2008; Carey & Xu, 2001; Spelke, 2000). Moreover, infants maintain their persisting object representations according to several spatiotemporal principles—for example, involving spatiotemporal continuity (e.g., Aguiar & Baillargeon, 1999, 2002; Spelke et al., 1995) and cohesion (Cheries et al., 2008; Chiang & Wynn, 2000; Huntley-Fenner et al., 2002; Wynn et al., 2000).

In the process of exploring the principles that control object persistence in infancy, researchers have also discovered striking examples of spatiotemporal priority. In a now-classic set of studies, Xu and colleagues explored the features that 10- and 12-month-old infants use to determine the number of objects hidden behind an occluder (Xu, 2002; Xu & Carey, 1996; Xu et al., 2004). This work revealed that 10-month-old infants reliably use spatiotemporal information to track the number of objects behind a screen. For example, they will reliably infer that two objects exist behind a screen on the basis of seeing them both at the same time, as in Fig. 6.7a. At the same time, however, infants of the same age will not reliably use contrasting surface features to draw this same inference. For example, they will not reliably infer that there are two objects behind the screen on the basis of seeing a truck and a duck sequentially emerge, as in Fig. 6.7b. Twelve-month-olds, in contrast, will use both sorts of information, similar to adults.

In a similar vein, 4-month-old infants have been shown to infer that two parts are in fact a single unitary occluded object (as shown in Fig. 6.1a) on the basis of spatiotemporal information (e.g., the fact that two parts separated by an occluder undergo common motion) but not featural information (e.g., the fact that two stationary parts separated by an occluder have similar colors and/or shapes; e.g., Kellman & Spelke, 1983; Kestenbaum et al., 1987). (Again, adults and older children will use both sorts of information.) Furthermore, some studies suggest that infants at these ages appear to use *only* spatiotemporal information to assess an object's unity: in the situation described earlier, for example, infants are perfectly happy to conceive of the two parts in common motion as a single object, despite the fact that their colors and shapes suggest strongly (to older children and adults) that they are distinct objects (but cf. Needham, 1997). Although the overall question of whether young infants *can* use surface features to fuel inferences about object persistence

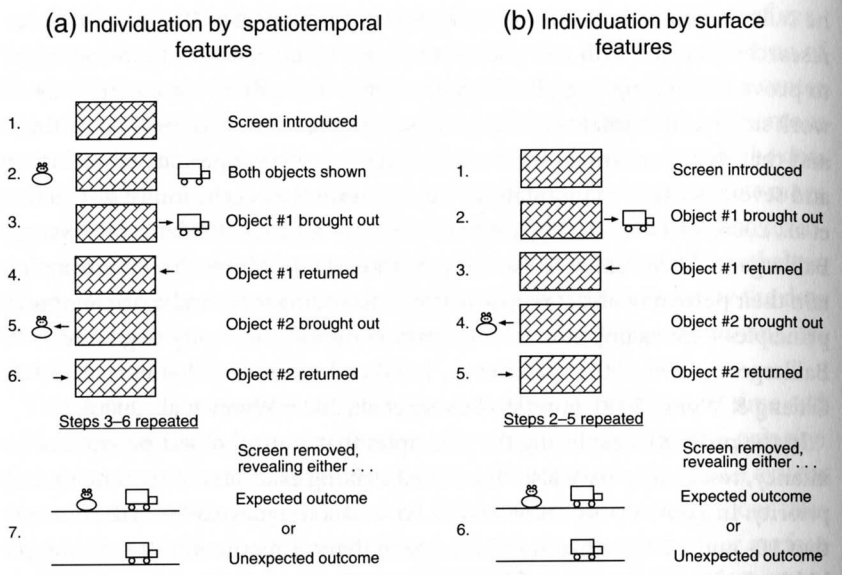

Fig. 6.7 Individuation in infancy. Examples of individuation by (a) spatiotemporal factors and (b) surface features in the seminal infancy studies of Xu and Carey (1996). (Adapted from Xu & Carey, 1996)

remains controversial (e.g., Leslie et al., 1998; Needham, 1998, 1999; Needham & Baillargeon, 1997, 1998; Needham et al., 2005; Wilcox, 1999; Wilcox & Baillargeon, 1998), this debate does not tar the overall conclusion of spatiotemporal priority. Overall, object persistence and the spatiotemporal rules that guide it appear to come online early in development, and these abilities may even reflect parts of 'core knowledge' that are a part of our genetic endowment (Spelke, 1988, 2000; Spelke & Kinzler, 2007).

6.3.2 The evolution of spatiotemporal priority

The developmental priority of spatiotemporal information raises the possibility that it may be evolutionarily prior as well. Over the past several years, comparative studies of other animals (typically, nonhuman primates) have demonstrated—perhaps unsurprisingly—that object persistence is a relatively ancient cognitive achievement. Using tasks similar to those used with human infants, researchers have discovered that several species of nonhuman primates represent objects as the same individuals over time and through occlusion. Similar to infants, they understand that objects continue to exist when occluded (e.g., chimpanzees: Beran, 2004; Beran & Beran, 2004; rhesus macaques: Hauser & Carey, 2003; Hauser et al., 1996, 2000; Santos et al., 2002; tamarins: Uller et al., 2001), retain their boundaries across motion

(e.g., rhesus macaques: Munakata et al., 2001), and trace continuous paths in space and time (e.g., rhesus monkeys: Santos, 2004; Santos & Hauser, 2002; tamarins: Santos et al., 2006).

Do nonhuman primates also prioritize spatiotemporal information when tracking persisting objects? We recently demonstrated that they do—and moreover that spatiotemporal priority can have powerful consequences for real-world behaviors such as foraging for food (Flombaum et al., 2004). In these experiments, rhesus macaques watched as a lemon rolled down a ramp and came to rest behind a tunnel (Occluder 1), and then as a kiwi emerged and became occluded at the end of its path behind a screen (Occluder 2; see Fig. 6.8).

When the kiwi emerged at about the time that the lemon should have (had it continued its motion), subjects searched for food only behind Occluder 2—apparently perceiving the lemon transform into a kiwi on the basis of spatiotemporally continuous motion. (Other control conditions verified that the monkeys did in fact recognize the featural difference.) In contrast, when a brief pause interrupted the occlusion of the lemon and the emergence of the kiwi, monkeys searched for food behind both occluders—apparently perceiving two distinct objects, and assuming (on the basis of the featural difference) that the lemon must have remained in the tunnel. In this way, the tunnel effect directly influenced not only the monkeys' percepts but also their subsequent spontaneous behaviors.

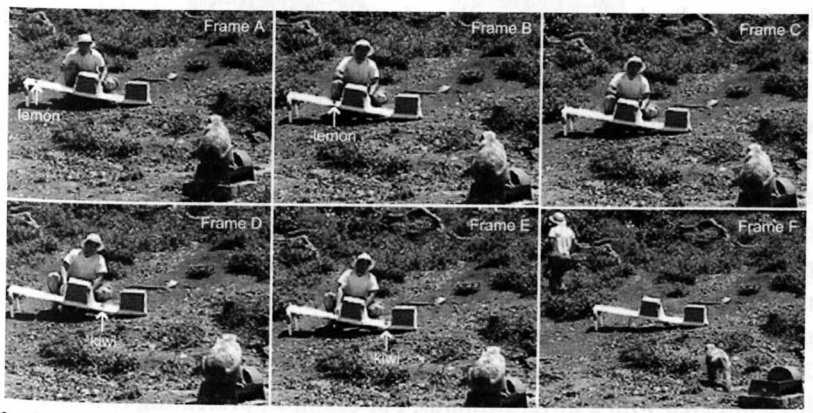

Fig. 6.8 The tunnel effect in rhesus macaques. Sample stills from a study of the tunnel effect in rhesus macaques (Flombaum et al., 2004). A monkey watches as a lemon rolls down a ramp (Frames A–B) and becomes concealed by Occluder 1 (Frame C), and then as kiwi rolls down the remainder of the ramp (Frames D–E) and becomes concealed by Occluder 2 (Frame F). Finally, the experimenter walks away, and the subject is given the opportunity to search (Frame F). (Adapted from Flombaum et al., 2004)

6.4 A neural signature of spatiotemporal cues to object persistence

The demonstrations of spatiotemporal priority discussed in the previous sections were drawn from many different kinds of behavioral studies. In contrast, the nature of spatiotemporal cues to object persistence has received much less study in cognitive neuroscience. Of course, the distinction between featural and spatiotemporal processing has been well characterized in neural terms. Perhaps most famously, these sorts of processes seem to be partially localized in anatomically distinct cortical streams (e.g., Livingstone & Hubel, 1988), with the ventral pathway corresponding to identification (i.e., the processing of *what* an object is, on the basis of surface features) and the dorsal pathway corresponding to individuation (i.e., how and *where* objects move, on the basis of spatiotemporal information). And a wealth of research in cognitive neuroscience has explored the neural bases of object identification on the basis of surface features. There have been few cognitive neuroscience studies of persistence as driven by spatiotemporal factors, however—and fewer still have studied both spatiotemporal properties and surface features at the same time. Several neuroscientific studies have demonstrated that neural processing of objects persists even when the objects are invisible, such as while momentarily occluded (e.g., Assad & Maunsell, 1995; Baker et al., 2001; Hulme & Zeki, 2007; Kaufman et al., 2005; Olson et al., 2003; Shuwairi et al., 2007). But few studies have examined the role of spatiotemporal principles such as continuity directly.

One recent exception is a functional magnetic resonance imaging (fMRI) study that compared spatiotemporal and featural processing (Yi et al., 2008). This study did not attempt to directly pit these types of processing against each other in general, as would be required in a comprehensive study of the neural bases of spatiotemporal priority. Instead, this study sought to explore how spatiotemporal factors might play a role even in some of the most staunchly 'featural' areas of visual cortex. We relied in this study on the phenomenon of neural *repetition attenuation* (e.g., Grill-Spector & Malach, 2001): a weaker neural response is typically observed to a repeated stimulus compared with a novel stimulus. For example, presenting two identical faces results in an attenuated fMRI response to the second face in the fusiform face area (FFA; an area of the visual cortex that responds selectively to faces), compared with the response to a second novel face. Thus, similar to the looking-time methods that are commonly employed in developmental and comparative studies (see Turk-Browne et al., 2008), repetition attenuation can help to reveal whether particular brain regions treat two

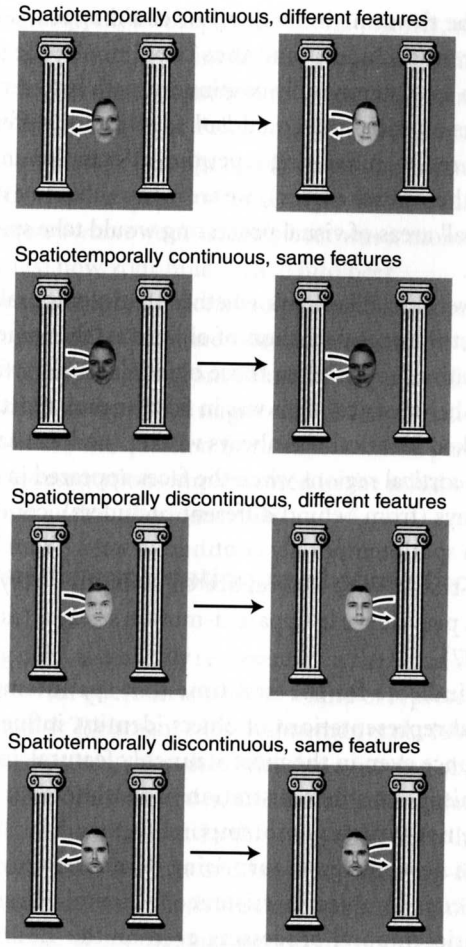

Fig. 6.9 Spatiotemporal continuity and surface features in an fMRI study. In an fMRI study of object persistence (Yi et al., 2008), faces could appear sequentially in a spatiotemporally continuous or discontinuous manner, and could look the same or different. Repetition attenuation in ventral cortical regions was weaker (or sometimes nonexistent) for repeated visual features in the spatiotemporally discontinuous condition, compared with the spatiotemporally continuous condition. (Dynamic animations are available online at http://www.yale.edu/perception/tunnel/.)

stimuli as two different objects or as instances of the same object—the key question in studies of persistence.

In our fMRI study (Yi et al., 2008), we presented two faces sequentially in a variety of situations. One situation—depicted in Fig. 6.9—involved dynamic occlusion: the faces could appear sequentially from behind the same occluder

(and thus could be the same individual) or could appear sequentially from behind two different occluders (and thus could not be the same individual, given the lack of a spatiotemporally continuous path between them). In addition, the two faces on each trial could *look* identical or different, in terms of their surface features. By measuring repetition attenuation in the FFA (and in other regions of the ventral cortex), we were thus able to ask about whether even these 'featural' areas of visual processing would take spatiotemporal factors into account.

In particular, we were able to ask whether spatiotemporal object discontinuity would affect the representation of objects as 'the same' (as revealed by repetition attenuation), even when those objects' visual surface features were (and appeared to be) identical. This was in fact the case: repetition attenuation for faces that looked identical was always weaker (and was sometimes nonexistent) in ventral cortical regions when the faces appeared in spatiotemporally discontinuous ways (from behind different occluders), compared with when they appeared in spatiotemporally continuous ways (from behind the same occluder). These results were also replicated in another very different type of display, based on persistence in apparent-motion streams (as in Fig. 6.5; Chun & Cavanagh, 1997).

These results indicate for the first time that spatiotemporal continuity modulates neural representations of object identity, influencing judgments of object persistence even in the most staunchly featural areas of the ventral visual cortex. This is not a demonstration of spatiotemporal priority per se, because we were not pitting spatiotemporal factors directly against featural factors. But, it is nevertheless a surprising demonstration of the power of spatiotemporal cues to object persistence in neural terms, as it shows the influence of spatiotemporal processing even on the 'home cortical turf' of featural processing. Indeed, the brain regions that exhibited these effects—the fusiform gyrus and lateral occipital cortex—are perhaps among the least likely places to find effects of spatiotemporal continuity, given their long association with the processing of what an object is on the basis of its surface features (e.g., Ungerleider & Mishkin, 1982). Thus, these results suggest that spatiotemporal object continuity is an important constraint on neural processing of identity in some of the best characterized regions of the visual cortex. This paves the way for future neuroscientific research to explore the existence of spatiotemporal priority per se.

6.5 Explaining spatiotemporal priority

The evidence reviewed in the preceding sections converges on spatiotemporal priority as a key principle of object persistence. Having now reviewed such a broad range of demonstrations of spatiotemporal priority, it is perhaps worth stepping back to again note how intuitively surprising it is. After all, when you reason using higher-level cognition through your own correspondence problems—such as finding your Prius in the crowded mall parking lot or determining which beverage belongs to whom on your dinner table—you do not exercise spatiotemporal priority! You might wonder just how your glass of milk got over *there* (where your friend's beer was a moment ago), but you wouldn't wonder just how your milk turned into beer.

Why, then, is spatiotemporal priority so powerful in more automatic processes of visual cognition? There are two ways to answer this question. In Section 6.5.2, we will approach it from a functional perspective by considering the computational advantages that might emerge from spatiotemporal priority. But first we consider the question from a mechanistic perspective: how are persisting object representations implemented in the mind so that spatiotemporal properties play such a dominant role?

6.5.1 A mechanistic explanation (and some thoughts on object files)

How are persisting objects mentally represented in the visual system? The most influential answer to this question appeals to a kind of representation called an *object file* (Kahneman & Treisman, 1984; Kahneman et al., 1992). An object file is a mid-level visual representation that 'sticks' to a moving object over time on the basis of spatiotemporal properties, and stores (and updates) information about that object's surface features. In this way, object files help to construct our conscious perception of objects and how they behave—for example, telling us 'which went where' or underlying the perception of object persistence despite featural change or momentary periods of occlusion. Critically, note that spatiotemporal and surface features are treated very differently in this framework. Object files can be thought of as analogous to 'file folders'—with one folder per salient object in the visual field, up to some limit of roughly 4 available folders—with addresses written on their covers that identify the object's current location and with various papers inside describing the object's features (perhaps including its past features). Spatiotemporal features can update a folder's label, perhaps even occasioning the creation of a new folder or the destruction of a current folder. Surface features, in contrast, are just along for the ride: when they change, the only result may be that the description stored inside the folder is modified.

Object files are thought to underlie object persistence via three steps (Kahneman et al., 1992): (a) a *correspondence* operation, which uses spatiotemporal information for each visual object to determine whether it is novel

or whether it moved from a previous location; (b) a *reviewing* operation, which retrieves previously stored object properties (e.g., color, shape) of those objects; and (c) an *impletion* operation, which uses both current and reviewed information to construct an evolving conscious percept, perhaps of object motion. Spatiotemporal priority arises in this model due to the assumption that only spatiotemporal information drives the correspondence operation, combined with the assumption that conflicting features are simply overwritten during the reviewing operation (although the previous features may also still be stored).

These three steps can account for a variety of the empirical demonstrations of spatiotemporal priority that were introduced in Sections 6.2 and 6.3. For example, the spatiotemporal rules that characterize amodal integration and apparent motion—such as the *nearest neighbor principle*, the *relative velocity principle*, and more generally, the attempt to minimize the distance and time interpolated—are presumably the very rules underlying the correspondence operation. Additionally, by the time that a change in surface features is processed in this framework (during the reviewing operation), decisions about object persistence have already been made (during the correspondence operation). As a result, property changes on their own cannot induce apparent motion (Navon, 1976), they have little or no impact on motion correspondences in apparent motion (Burt & Sperling, 1981), and they do not preclude amodal integration as in the tunnel effect (Burke, 1952; Flombaum & Scholl, 2006). Finally, illusory conjunctions seem less mysterious in this context: when the impletion operation makes an error while building a percept, it can only make the error by misbinding ('misfiling') one of the surface properties collected during the reviewing procedure; the object's locations, in contrast, have already been determined by the correspondence operation. In many ways, the object representation *is* the object's location—such that the file can only be addressed via its spatiotemporal label—which is perhaps why surface features can be illusorily bound to the wrong locations, but not vice versa.

What evidence is there that object files exist and work in these ways? Object files are typically studied with a paradigm known as object reviewing, depicted in Fig. 6.10a. In the initial preview display, a number of distinct objects are presented and then letters (or other symbols) briefly appear on some or all of them. After the letters disappear, the objects begin moving about the display. Once they come to rest, a single target letter appears on one of the objects, and the observer's task is simply to name that letter as quickly as possible. This response is typically slightly faster when the letter matches one of the initially presented letters (a type of display-wide priming). However, observers are faster still to name the target letter when it is the same letter that initially appeared on that same object (in a congruent trial), compared with when the

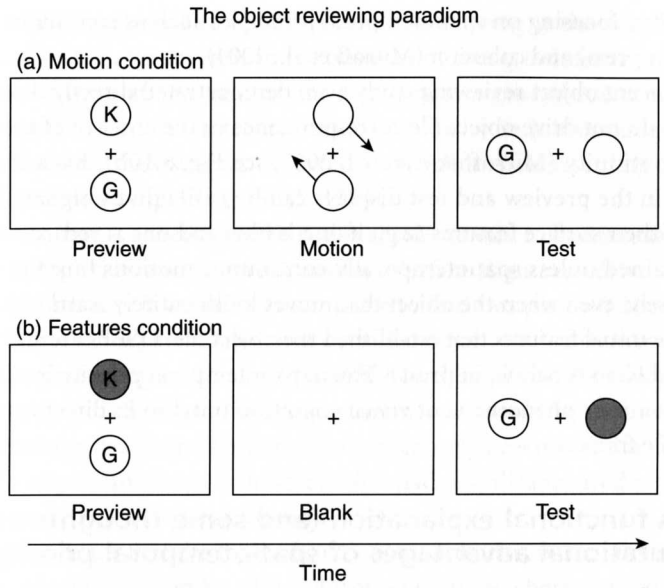

Fig. 6.10 Object reviewing. (a) In the standard 'object reviewing paradigm' (Kahneman et al., 1992), observers first see two objects with a letter presented in each one (the preview screen). Next, the letters disappear and the two objects move to new positions. Finally, the objects stop moving, a letter appears in one of the objects (the test screen), and observers must indicate whether the letter had appeared in the preview screen. A response-time advantage (called an 'object-specific preview benefit', or OSPB) is obtained when the test letter appears in the same object as it had in the preview screen, compared with when it appears in a different object. (b) As demonstrated by Mitroff and Alvarez (2007), however, no OSPB occurs when shared surface features (e.g., color) link the objects in the preview and test screens, without continuous motion.

final letter initially appeared on a different object (in an incongruent trial). This is termed an *object-specific preview benefit* (OSPB). This relative response-time advantage for congruent trials over incongruent trials is necessarily object-based because the objects' spatial locations change from the preview to the target displays.

Several more recent object reviewing studies have employed a different method wherein the task is to answer as quickly as possible whether the target letter matches *either* of the initial previews. This forces subjects to attend to the preview information (which otherwise can be completely ignored), and perhaps for that reason it tends to produce more robust OSPBs (Kruschke & Fragassi, 1996; Noles et al., 2005). Object reviewing studies of this sort have recently been used to explore the rules that govern the creation and maintenance of

object files, focusing on spatiotemporal principles such as continuity (Gao & Scholl, in press) and cohesion (Mitroff et al., 2004).

One recent object reviewing study even demonstrated directly that surface features do not drive object file correspondence in the absence of spatiotemporal continuity (Mitroff & Alvarez, 2007; see Fig. 6.10b). Even if the two objects in the preview and test displays can be easily disambiguated on the basis of their surface features (e.g., if one is blue and one is red), an OSPB is not obtained unless spatiotemporally continuous motions link the objects. Conversely, even when the object that moves looks entirely featurally distinct from the initial features that established the object file, OSPBs are still reliably obtained (Gao & Scholl, in press). Thus, spatiotemporal priority is salient not only in various phenomena of visual cognition but also in direct tests of the object file framework.

6.5.2 A functional explanation (and some thoughts on the computational advantages of spatiotemporal priority)

Object file theory provides an elegant explanation for how spatiotemporal priority is implemented, but it does not explain *why* the visual system would be constructed in this way. We suggest, however, that there are at least two reasons for why spatiotemporal priority is an adaptive solution for the kinds of correspondence problems typically faced by the visual system.

6.5.2.1 Avoiding coincidences

The first reason why spatiotemporal priority might be adaptive is that, statistically speaking, spatiotemporal priority means that the visual system rejects coincidences—and coincidences, by definition, are statistically improbable, and thus *should* be avoided.

On the face of it, ignoring surface features altogether can seem shortsighted. In our experiments on the tunnel effect with monkeys, for example, a kiwi enters an occluder and then a lemon exits on the other side, and because of spatiotemporal continuity monkeys perceive a single enduring object (Flombaum et al., 2004). Why does their visual system ignore the fact that the two objects look so different, leading them to search for only a single piece of food? (After all, this means that they get less food!)

Although it is true that featural evidence is being ignored in the tunnel effect, it is also true that the spatiotemporal evidence is statistically more reliable. The particular spatiotemporal evidence that drives the tunnel effect is the continuous motion that links the occluding and disoccluding objects. If the events on either side of the occluder reflected two distinct objects, then this continuity would be unexplained: in other words, it would just be a coincidence that the

second object happened to appear when (and moving at the speed) it did. The visual system, however, is in the business of trying to infer the structure of the world from utterly insufficient data, and as a general rule it would be unwise for it to settle on interpretations that are by definition judged to be unlikely. As such, the visual system avoids coincidences when settling on interpretations of visual input (e.g., Rock, 1983).

This reasoning is general, and so it also applies to other phenomena such as amodal completion and apparent motion. Perfect alignment drives amodal completion because it would be improbable to find two perfectly aligned segments that were not parts of the same extended object. (If you were to toss two physical segments up in the air, what is the likelihood that they would land perfectly aligned?) Similarly, in apparent motion, it would be an unexplained coincidence for the second flash to reflect an entirely new object that just happened to appear so close to an object that had just disappeared a moment before—and thus, we instead see motion.

In fact, evidence that the logic of spatiotemporal priority has something to do with coincidences can be adduced from one study where surface features have actually been shown to affect object persistence. In the *flash-lag illusion*, observers misperceive the spatial relationship between a moving object and a briefly flashed object (MacKay, 1958; Nijhawan, 2002). The illusion is not obtained, however, if the moving object stops at exactly the moment that the flash takes place. In other words, the object's trajectory must persist following the flash to induce the misperception. The illusion also is not obtained if the moving object changes its size or shape—but still continues to move—at the very moment that the second object is flashed (Moore & Enns, 2004; see also Moore et al., 2007).

Putting these facts together suggests that the sudden featural change induces the individuation of a new object, thus fracturing the motion trajectories before and after the flash. This is atypical: here a featural change seems to cause the creation of a new object, rather than just updating the features of an existing object representation. This special case can be reconciled with spatiotemporal priority, however, in terms of coincidences: here the change in surface features is occurring at a particular time—at the exact moment of the sudden, unexpected flash—and this coincidence is 'explained' by the visual system in terms of the flash and the feature change both reflecting the appearance of a new object[1].

[1] There is at least one other way that this type of study (see also Moore et al., 2007) may be reconciled with spatiotemporal priority. The particular 'surface features' used here (size and shape) may in fact be better characterized as spatial features of a sort—part of the

6.5.2.2 Surface features are unreliable, but spatiotemporal rules can be implemented without exceptions

The second reason why spatiotemporal priority might be adaptive is simply that surface features are often more unreliable guides to persistence than are spatiotemporal cues. The former must be indirectly inferred, and an instance of a changing surface feature may or may not reliably signal a new object. In contrast, some spatiotemporal rules for persistence can be computationally implemented without exception, as extremely reliable guides to persistence.

In the first place, it is obvious that surface features alone could never be a sole guide to object persistence. If they were, then we could never keep track of multiple individuals that looked the same. This would render certain types of real-world hunting and predation impossible, because when chasing a herd of animals it may often be impossible in practice to discriminate individual prey on the basis of their surface features. We can obviously still keep track of individuals in such contexts, though, and this is also possible in the laboratory. In multiple object tracking (MOT) experiments, for example, observers must track a group of haphazardly and independently moving targets among a group of moving distractors, and this is possible (for up to 4 targets) even when all of the objects look identical (Pylyshyn & Storm, 1988; for a review, see Scholl, in press). Indeed, recent experiments have demonstrated that MOT seems to rely solely on spatiotemporal cues in some ways: for example, additional moving objects in the display will be treated as distractors even if they have completely different surface properties, but additional static objects will not be treated as distractors even if they are identical to the targets (Pylyshyn et al., in press).

Even when distinctive surface features are present in a display, however, they still may not be reliable guides to identity over time. In part, this is because many objects can change their surface properties over time, although not becoming new individuals: chameleons and octopuses can change their colors from moment to moment, for example, and even humans can radically change their postures and facial expressions, and so on.

In addition, however, many surface features cannot be directly recovered from retinal input, but must be inferred indirectly. We think of lightness, for example, as a surface feature of objects, but what we actually need to perceive

global spatial signature of *where* an object is. It is seductive to think of locations as single points, but in fact the spatiotemporally encoded locations of objects may have spatial extent and even a rough type of overall boundary shape (cf. Alvarez & Cavanagh, 2008). We are exploring this possibility in current research.

in order for lightness to be useful is the *reflectance* of the object. Recovering reflectance is technically impossible to reliably infer, however, because the actual number of photons reaching the eyes after reflecting off a surface will always be the product of the illumination from the light source(s) (which we want to discount) and the reflectance of the surface (the 'paint', which we want to know about). Separating these sources is not strictly possible, as it would essentially require the visual system to 'unmultiply', for example, solving for R in the equation $R \times I = 12$ (see Adelson, 2000). Instead, the visual system must infer reflectance indirectly using a set of heuristic assumptions. These indirect inferences typically succeed in everyday life, of course, but in general the visual system has no way of checking the reliability of such inference in individual cases.

In contrast, some spatiotemporal properties such as relative position can be much more reliably recovered from retinal input, because relative separation on the retina directly entails relative separation in the world. And, when computations of such features are recovered, they are much more reliable cues to persistence. If two objects are encountered in distinct locations without a possible spatiotemporally continuous path between them, for example, then they cannot be the same individual—no matter how similar they might appear. (And as we saw in Section 6.4, even those areas of the visual cortex that are most involved in the processing of surface features appear to be sensitive to this fact.) Indeed, this can be a direct reflection of physical laws, whereas surface features can at best be merely strongly suggestive of persistence or the lack thereof. As a result, the reliability of such spatiotemporal cues to persistence may also give them a computational advantage: they can be implemented in the mind *without exception*.

6.6 Conclusion

The principle of spatiotemporal priority, as we have seen, has considerable scope in accounting for phenomena of object persistence. It describes a general solution to types of correspondence problems as they are grappled with in many guises in visual cognition research—including the tunnel effect, apparent motion, illusory conjunctions, and object reviewing. It captures some of the primary features of the dominant 'object file' model of object persistence. It accounts for the perception of persisting objects not only in adult humans but also in infants and nonhuman primates. And it may have a firm foundation in both mechanistic and functional terms. We suggest, as captured in the title of this chapter, that it constitutes a fundamental principle of object persistence.

References

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*, 2nd edn. (pp. 339–351). Cambridge, MA: MIT Press.

Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology, 39*, 116–157.

Aguiar, A., & Baillargeon, R. (2002). Developments in young infants' reasoning about occluded objects. *Cognitive Psychology, 45*, 267–336.

Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Perception and Psychophysics, 70*, 346–364.

Anstis, S. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London B, 290*, 153–168.

Assad, J., & Maunsell, J. (1995) Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature, 373*, 518–521.

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: object permanence in 6- and 8-month old infants. *Cognition, 23*, 21–41.

Baillargeon, R. (2008). Innate ideas revisited: for a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science, 3*, 2–13.

Baker, C. I., Keysers, C., Jellema, T., Wicker, B., & Perrett, D. I. (2001). Neuronal representation of disappearing and hidden objects in temporal cortex of the macaque. *Experimental Brain Research, 140*, 375–381.

Beran, M. J. (2004). Chimpanzees (*Pan troglodytes*) respond to nonvisible sets after one-by-one addition and removal of items. *Journal of Comparative Psychology, 118*, 25–36.

Beran, M. J., & Beran, M. M. (2004) Chimpanzees remember the results of one-by-one addition of food items to sets over extended time periods. *Psychological Science, 15*, 94–99.

Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology, 4*, 121–138.

Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review, 88*, 171–195.

Carey, S., & Xu, F. (2001). Infant knowledge of objects: Beyond object files and object tracking. *Cognition, 80*, 179–213.

Cheries, E. W., Mitroff, S. R., Wynn, K., & Scholl, B. J. (2008). Cohesion as a principle of object persistence in infancy. *Developmental Science, 11*, 427–432.

Cheries, E. W., Wynn, K., & Scholl, B. J. (2006). Interrupting infants' persisting object representations: An object-based limit? *Developmental Science, 9*, F50–F58.

Chiang, W-C., & Wynn, K. (2000). Infants' representation and tracking of multiple objects. *Cognition, 77*, 169–195.

Chun, M. M., & Cavanagh, P. (1997). Seeing two as one: linking apparent motion and repetition blindness. *Psychological Science, 8*, 74–79.

Dawson, M. (1991). The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem. *Psychological Review, 98*, 569–603.

Flombaum, J. I., Kundey, S., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in rhesus macaques: a study of the tunnel effect. *Psychological Science, 15*, 795–800.

Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 840–853.

Gao, T., & Scholl, B. J. (in press). Are objects required for object-files? Roles of segmentation and spatiotemporal continuity in computing object persistence. *Visual Cognition*.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica, 107*, 293–321.

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: examinations of content and format. *Cognitive Psychology, 47*, 367–401.

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London: Biological Science, 267*, 829–833.

Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Science, 93*, 1514–1517.

Hulme, O. J., & Zeki, S. (2007). The sightless view: neural correlates of occluded objects. *Cerebral Cortex, 17*, 1197–1205.

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A., & Klempen, N. (2000). Fractionating the binding process: Neuropsychological evidence distinguishing binding of form from binding of features. *Vision Research, 40*, 1569–1596.

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: perceived rigidity and cohesiveness influence infants' representations of small groups of discrete entities. *Cognition, 85*, 203–221.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, & D. R. Davies (Eds.), *Varieties of Attention*. New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kanizsa, G. (1979). *Organization in Vision*. New York: Praeger.

Kanizsa, G. (1982). Amodal completion: Seeing or thinking? In J. Beck (Ed.), *Organisation and Representation in Perception* (pp. 167–190). New Jersey: Erlbaum.

Kanwisher, N. G. (1987). Repetition blindness: type recognition without token individuation. *Cognition, 27*, 117–143.

Kaufman, J., Csibra, G., & Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proceedings of the National Academy of Sciences, 102*, 15271–15274.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23*, 141–221.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*, 483–524.

Kestenbaum, R., Termine, N., & Spelke, E. S. (1987). Perception of objects and object boundaries by three-month-old infants. *British Journal of Developmental Psychology, 5*, 367–383.

Kolers, E. (1972). *Aspects of Motion Perception*. New York: Pergamon Press.

Kolers, P., & Pomerantz, J. (1971). Figural change in apparent motion. *Journal of Experimental Psychology, 87*, 99–108.

Kruschke, J. K., & Fragassi, M. M. (1996). The perception of causality: feature binding in interacting objects. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 441–446). Hillsdale, NJ: Erlbaum.

Leslie, A. M., Xu, F., Tremoulet, P., & Scholl, B. J. (1998). Indexing and the object concept: developing 'what' and 'where' systems. *Trends in Cognitive Sciences, 2*, 10–18.

Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, 240*, 740–749.

MacKay, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature, 181*, 507–508.

Michotte, A., Thinès, G., & Crabbé, G. (1964/1991). Les complements amodaux des structures perceptives. In *Studia Psychologica*. Louvain: Publications Universitaires. Reprinted and translated as Michotte, A., Thinès, G., & Crabbé, G. (1991). Amodal completion of perceptual structures. In G. Thines, A. Costall, & G. Butterworth (Eds.) *Michotte's Experimental Phenomenology of Perception* (pp. 140–167). Hillsdale, NJ: Erlbaum.

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, underlie object persistence. *Psychonomic Bulletin and Review, 14*, 1199–1204.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2004). Divide and conquer: how object files adapt when a persisting object splits into two. *Psychological Science, 15*, 420–425.

Moore, C. N., & Enns, J. T. (2004). Object updating and the flash-lag effect. *Psychological Science, 15*, 866–871.

Moore, C. M., Mordkoff, J. T, & Enns, J. T. (2007). The path of least persistence: evidence of object-mediated visual updating. *Vision Research, 47*, 1624–1630.

Navon, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 130–138.

Needham, A. (1997). Factors affecting infants' use of featural information in object segregation. *Current Directions in Psychological Science, 6*, 26–33.

Needham, A. (1998). Infants' use of featural information in the segregation of stationary objects. *Infant Behavior and Development, 21*, 47–76.

Needham, A. (1999). How infants grasp two adjacent objects: effects of perceived display composition on infants' actions. *Developmental Science, 2*, 219–233.

Needham, A., & Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition, 62*, 121–149.

Needham, A., & Baillargeon, R. (1998). Effects of prior experience in 4.5-month-old infants' object segregation. *Infant Behavior and Development, 21*, 1–24.

Needham, A., Dueker, G. L., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition, 94*, 215–240.

Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect. *Trends in Cognitive Sciences, 6*, 387–393.

Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception and Psychophysics, 67*, 324–334.

Olson, I. R., Gatenby, J. C., Leung, H. C., Skudlarski, P., & Gore, J. C. (2003). Neuronal representation of occluded objects in the human brain. *Neuropsychologia, 42*, 95–104.

Piaget, J. (1954). *The construction of reality in the child*. **New York: Basic Books.**

Pylyshyn, Z. W., Haladjian, H., King, C., & Reilly, J. (in press). Selective nontarget inhibition in multiple object tracking (MOT). *Visual Cognition.*

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179–197.

Rafal, R. D. (1997). Balint syndrome. In T. Feinberg, & M. Farah (Eds.), *Behavioral Neurology and Neuropsychology* (pp. 337–356). New York: McGraw-Hill.

Ramachandran, V. S., & Anstis, S. M. (1986). The perception of apparent motion. *Scientific American, 245*, 102–109.

Robertson, L., Treisman, A., Freidman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways: evidence from Balint's syndrome. *Journal of Cognitive Neuroscience, 9*, 254–276.

Rock, I. (1983). *The Logic of Perception*. Cambridge, MA: MIT Press.

Santos, L. R. (2004). 'Core knowledges': a dissociation between spatiotemporal knowledge and contact-mechanics in a non-human primate? *Developmental Science, 7*, 167–174.

Santos, L. R., & Hauser, M. D. (2002). Dissociations between implicit knowledge and search behavior in cotton-top tamarins. *Developmental Science, 5*, F1–F7.

Santos, L. R., Seelig, D., & Hauser, M. D. (2006). Cotton-top tamarins' (*Saguinus oedipus*) expectations about occluded objects: a dissociation between looking and reaching tasks. *Infancy, 9*, 147–171.

Santos, L. R., Sulkowski, G., Spaepen, G. M., & Hauser, M. D. (2002). Object individuation using property/kind information in rhesus macaques (*Macaca mulatta*). *Cognition, 83*, 241–264.

Scholl, B. J. (2001b). Spatiotemporal priority and object identity. *Cahiers de Psychologie Cognitive, 20*, 359–371.

Scholl, B. J. (in press). What have we learned about attention from multiple object tracking (and vice versa)? In D. Dedrick, & L. Trick (Eds.), *Computation, Cognition, and Pylyshyn*. Cambridge, MA: MIT Press.

Scholl, B. J., & Flombaum, J. I. (in press). Object persistence. In B. Goldstein (Ed.), Encyclopedia of Perception. Thousand Oaks, CA: Sage Publications.

Shechter, S., Hochstein, S., & Hillman, P. (1988). Shape similarity and distance disparity as apparent motion correspondence cues. *Vision Research, 28*, 1013–1021.

Shuwairi, S. M., Curtis, C. E., & Johnson, S. P. (2007). Neural substrates of dynamic object occlusion. *Journal of Cognitive Neuroscience, 19*, 1275–1285.

Singh, M., & Fulvio, J. (2005). Visual extrapolation of contour geometry. *Proceedings of the National Academy of Sciences, 102*, 939–944.

Spelke, E. (1988). The origins of physical knowledge. In L. Weiskrantz (Ed.), *Thought without language* (pp. 168–184). Oxford: Oxford Science Publications.

Spelke, E.S. (2000). Core knowledge. *American Psychologist, 55*, 1233–1243.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*, 605–632.

Spelke, E. S., Kestenbaum, R., Simons, D., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *The British Journal of Developmental Psychology, 13*, 113–142.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*, 89–96.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107–141.

Turk-Browne, N. B., Scholl, B. J., & Chun, M. M. (2008). Babies and brains: habituation in infant cognition and functional neuroimaging. *Frontiers in Human Neuroscience, 2*, 16. doi: 10.3389/neuro.09.016.2008.

Uller, C., Hauser, M., & Carey, S. (2001). The spontaneous representation of number in a New World primate species, cotton-top tamarins. *Journal of Comparative Psychology, 115*, 1–10.

Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. Mansfield (Eds.), *Analysis of Visual Behaviour* (pp. 549–586). Cambridge, MA: MIT Press.

Walker, P., & Davies, S. J. (2003). Perceptual completion and object-based representations in short-term visual memory. *Memory and Cognition, 31*, 746–760.

Wertheimer, M. (1912/1961). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie, 61*,161–265. Reprinted and translated as Wertheimer, M. (1961). Experimental studies on the seeing of motion. In T. Shipley (Ed. and Trans.) *Classics in Psychology* (pp. 1032–1089). New York: Philosophical Library.

Wilcox, T. (1999). Object individuation: infants use of shape, size, pattern, and color. *Cognition, 72*, 125–166.

Wilcox, T., & Baillargeon, R. (1998). Object individuation in infancy: the use of featural information in reasoning about occlusion events. *Cognitive Psychology, 37*, 97–155.

Wu, D. A., Kanai, R., & Shimojo, S. (2004). Steady-state misbinding of colour and motion. *Nature, 429*, 262.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749–750.

Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition, 83*, B55–B62.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85*, 223–250.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology, 30*, 111–153.

Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology, 49*, 155–190.

Yi, D-J., Turk-Browne, N. B., Flombaum, J. I., Kim, M-S., Scholl, B. J., & Chun, M. M. (2008). Spatiotemporal object continuity in human ventral visual cortex. *Proceedings of the National Academy of Sciences, 105*, 8840–8845.

Chapter 7

Infants' representations of material entities

Rebecca D. Rosenberg & Susan Carey

7.1 Infants' representations of material entities

Although the core cognition hypothesis is controversial, many researchers endorse the existence of systems of core cognition. Hypothesized domains of core cognition include objects and physical causality, agents, and number (e.g., Baillargeon, 2001; Carey, in press; Carey & Spelke, 1994, 1996; Leslie, 1994; Spelke & Kinzler, 2007). By hypothesis, systems of core cognition exhibit several properties that distinguish the representations within them from other types of mental representations. These distinctive properties include (1) conceptual content, (2) innate input analyzers that pick out entities in core domains and constrain representations of the events these entities participate in, (3) continuity through the lifespan, (4) domain-specific learning mechanisms, and (5) modularity.

The first of these properties differentiate the representations in core cognition from perceptual representations; the content of the symbols in core cognition go beyond the vocabulary of sensory and perceptual primitives and the outputs of core cognition participate in central inferential processes. The remaining properties differentiate core cognition from most human conceptual representations. There are no innate input analyzers that identify most of the entities in the world we think about (from electrons to ghosts and Gods to hammers), most concepts are vulnerable to change, even abandonment, and although intuitive theories provide constraints on the information relevant to confirmation of hypotheses within them, these processes are neither modular nor encapsulated (Fodor, 1983).

Cognition about objects was the first system of core cognition to be studied (e.g., Baillargeon, 1986; Spelke et al., 1992) and it remains a parade case. Object representations go beyond the vocabulary of sensorimotor primitives and in this sense are conceptual. Bounded, coherent, independently movable entities are represented as spatiotemporally continuous and as continuing to exist when out of sight, and representations of their motion are subject to the