Journal of Experimental Psychology: General
2017, Vol. 146, No. 3, 362–375

0096-3445/17/$12.00 http://dx.doi.org/10.1037/xge0000270

Exploiting Core Knowledge for Visual Object Recognition

Mark W. Schurgin and Jonathan I. Flombaum
Johns Hopkins University

Humans recognize thousands of objects, and with relative tolerance to variable retinal inputs. The acquisition of this ability is not fully understood, and it remains an area in which artificial systems have yet to surpass people. We sought to investigate the memory process that supports object recognition. Specifically, we investigated the association of inputs that co-occur over short periods of time. We tested the hypothesis that human perception exploits expectations about object kinematics to limit the scope of association to inputs that are likely to have the same token as a source. In several experiments we exposed participants to images of objects, and we then tested recognition sensitivity. Using motion, we manipulated whether successive encounters with an image took place through kinematics that implied the same or a different token as the source of those encounters. Images were injected with noise, or shown at varying orientations, and we included 2 manipulations of motion kinematics. Across all experiments, memory performance was better for images that had been previously encountered with kinematics that implied a single token. A model-based analysis similarly showed greater memory strength when images were shown via kinematics that implied a single token. These results suggest that constraints from physics are built into the mechanisms that support memory about objects. Such constraints—often characterized as 'Core Knowledge'—are known to support perception and cognition broadly, even in young infants. But they have never been considered as a mechanism for memory with respect to recognition.

Keywords: Core Knowledge, long-term memory, object recognition, spatiotemporal continuity, tolerance

The problem of object recognition—whether accomplished by a person, animal, or computer system—is to recognize objects in the present on the basis of experience in the past. It is a computational problem because changes in viewpoint, orientation, and lighting conditions (among other things) can make the same object look different across encounters (Cox & Dicarlo, 2008; Cox, Meier, Oertelt, & DiCarlo, 2005; DiCarlo & Cox, 2007; Rust & Stocker, 2010; Wallis & Bulthoff, 2001). In other words, an object can look very different from itself depending on how and when it is viewed, and this renders pixel-wise image comparisons an inadequate strategy for recognition (DiCarlo, Zoccolan, & Rust, 2012; Logothetis & Sheinberg, 1996).

Yet humans possess relatively invariant and impressive recognition abilities, seemingly able to recognize thousands of objects (Biederman, 1987; Shepard, 1967; Standing, Conezio, & Haber, 1970) following minimal exposure (Potter, 1976; Thorpe, Fize, & Marlot, 1996) and despite changes to input properties. It remains unknown exactly how this is accomplished, and this remains one area in which artificial systems have yet to surpass humans (Andreopoulos & Tsotsos, 2013; DiCarlo, Zoccolan, & Rust, 2012; Pinto, Cox, & DiCarlo, 2008).

One known piece of the process involves a temporal association rule (Isik, Leibo, & Poggio, 2012). Over the short-term an experience that may otherwise be thought of as a single encounter with an object—perhaps lasting only a few seconds—may be better conceived as a collection of noisy encounters with varying input quality and structure. This creates an opportunity for learning. Cataloguing the ways that object-related inputs change in the short-term supplies a basis for knowing the type of variability to expect from an object in the future. Temporal association of this kind has been shown to support object recognition (Cox et al., 2005; Wallis, 2002; Wallis & Bulthoff, 2001) and to accommodate explicit computational algorithms (Isik et al., 2012).

But like many simple strategies, temporal association faces a frame problem (Dennett, 2006; Fodor, 1983). Merely correlating inputs that arrive in close succession will lead to the mixing of signals from *different* sources. Temporal association therefore requires a means to ensure that only signals from shared sources become associated during the process of encoding. One such means is to exploit eye movements. If an observer sees an object in her periphery and then saccades to foveate the object, she can learn by associating the two inputs. Psychophysical, neural, and computational evidence have shown that temporal association sup-

Mark W. Schurgin and Jonathan I. Flombaum, Department of Psychological and Brain Sciences, Johns Hopkins University.

We thank Patricia Kingkeo, Victor Kang, and Gustavo Beruman for assistance with data collection, Hee Yeon Im for assistance with image noise programming, Zheng Ma and Colin Wilson for modeling expertise and advice, and Jack Schurgin for thoughtful commentary. This research was funded by NSF BCS-1534568 and a seed grant from the Johns Hopkins University Science of Learning Institute to Jonathan I. Flombaum A pilot experiment related to but not present in the current manuscript was published in *Visual Cognition* in 2013 (Schurgin et al., 2013). Some of the results and ideas in this article were also presented at the annual meeting on Object Perception, Attention, and Memory (OPAM) in 2013, and at the annual meeting of the Visual Sciences Society (VSS) in 2014 and 2015.

Correspondence concerning this article should be addressed to Mark W. Schurgin, Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 North Charles Street, Ames Hall 127, Baltimore, MD 21218. E-mail: maschurgin@jhu.edu

ports object memory through saccades, even demonstrating that invariant memory can be 'broken'—led astray—if an experimenter changes stimuli surreptitiously during the milliseconds between a saccade onset and termination (Cox et al., 2005; Isik et al., 2012; Li & DiCarlo, 2008; Poth, Herwig, & Schneider, 2015; Poth & Schneider, 2016). In these situations, an observer is tricked; but in the real-world the strategy should be reliable nearly all the time. The target of a saccade and its terminus will usually be the same object.

We sought to investigate a second and unconsidered strategy for temporal association: that the visual system exploits the rules of object kinematics to limit the scope of association to inputs with the same individual object as their source. By 'the rules of object kinematics' we mean constraints from physics that can be used to determine when an object seen at one moment in time is likely to be the same *individual* as an object seen later. These are often referred to as the rules of spatiotemporal persistence (Scholl & Flombaum, 2010). Note that by 'the same individual' we mean 'the same exact exemplar' (or token), as opposed to 'an example of the same thing or kind of thing.'

Expectations about spatiotemporal persistence appear to support a great deal of visual and cognitive processing. They are known to constrain how children and infants evaluate and learn about the physical world (Baillargeon, Spelke, & Wasserman, 1985; Spelke, Kestenbaum, Simons, & Wein, 1995; Stahl & Feigenson, 2015; Wilcox, & Baillargeon, 1998a; Wilcox, & Baillargeon, 1998b; Xu & Carey, 1996). They also play a critical role in the motivation and theorizing underlying influential theories of midlevel object representation (Kahneman, Treisman, & Gibbs, 1992). And they are known to influence the processing, categorization, and working memory representation of objects over short periods of time (on the order of seconds; Flombaum & Scholl, 2006; Flombaum, Scholl, & Santos, 2009; Yi et al., 2008). More generally, these mechanisms are thought to be either innate or to arise early in human development, and preserved through evolutionary history for their obvious utility (Spelke & Kinzler, 2007). At least one previous study found effects of spatiotemporal understanding on the foraging behavior of a nonhuman species, the Rhesus monkey (Flombaum, Kundey, Santos, & Scholl, 2004).

Yet expectations about spatiotemporal persistence have been scarcely considered as a mechanism of long-term learning (but see Stahl & Feigenson, 2015), particularly with respect to the challenge of object recognition. We hypothesized that the rules of spatiotemporal persistence should constrain long-term object recognition.

To investigate the hypothesis, we manipulated perceived object persistence by manipulating motion continuity. Object perception naturally involves attributions of token identity, attributions of 'which' as opposed to 'what.' As a tossed ball passes by, one perceives the ball's physical properties, and also that it is the *same* ball from moment to moment. This is despite the fact that the ball's projection on the retina mutates considerably over the ball's journey. Conversely, if you see two identical people walking past each other, the natural inference is that they are twins (because the same object cannot be in two places at once). This is despite the (nearly) identical physical appearances of the individuals.

To manipulate perceived token identity, we generated animations in which two objects moved in opposite directions and passed one another. The animations relied on apparent motion (Anstis & Mackay, 1980; Chun & Cavanagh, 1997; Yi et al., 2008). In apparent motion, noncontiguous visual transients are perceived as instances of a single object moving continuously, provided that the transients occur close enough to one another in time and space. We will use the term 'stream' to refer to a perceived continuous motion path linking any number of transients. In our main experiment, each trial included two streams moving in opposite directions from one side of the display to another (Figure 1; dynamic demonstrations can be viewed online at jhuvisualthinkinglab.com/demos-schurgin-and-flombaum-stcontinuity). Each transient in a stream was usually the onset and then offset of a collage-like mask within a rectangular boundary (we will call these 'mask-transients'). But two transients in each display were the onset and offset of an image of a real-world object (we will call these 'image-transients').

Within a trial, the two image-transients were always images of the same object. In half of the trials, the image-transients appeared in positions within the same stream (supplanting the mask-transients that would have appeared in those two positions otherwise). Thus, the images would be perceived as instances of the same token in two different places at different moments. We call this 'continuous motion.' In 'discontinuous motion' trials, the image-transients appeared sequentially in different streams, so that

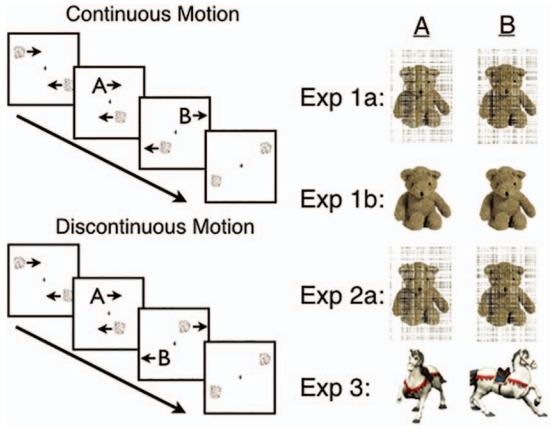

Figure 1. Procedure of the incidental encoding task. Each trial was made up of four frames, producing two apparent motion streams moving in opposite directions. Each stream comprised random noise masks (mask-transients) at their most eccentric initial and final positions during the first and last frames of a trial. [See Chun & Cavanagh, 1997, upon which these methods were closely modeled; see also Yi et al., 2008]. In the positions closer to fixation, two mask-transients were replaced by two appearances of an image (image-transients) in succession (i.e., in the second and the third frame of a trial). The critical manipulation was whether the successive image appearances supplanted mask-transients in a single stream (continuous motion) or in different streams (discontinuous motion). In the figure, A and B designate the positions in which image-transients supplanted mask-transients for each of the two motion conditions. In Experiment 1a and Experiment 2a the image-transients were photos of an object embedded in independent noise during each of its appearances. In Experiment 1b the photos were noiseless (during encoding). In Experiment 3 the photos were of an object at one orientation and then of the same object at a different orientation. In this experiment the recognition test always showed a third, previously not shown orientation. See the online article for the color version of this figure.

despite their similar (physical) appearances, they would be perceived as instances of two different tokens.

Exposure to images took place during an incidental encoding paradigm, typical of research on long-term memory (Blumenfeld & Ranganath, 2006; Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Kirchhoff, Wagner, Maril, & Stern, 2000; Kim & Yassa, 2013; Wittmann et al., 2005). Each trial included two motion streams, as just described. The repeated image was unique in each trial. We subsequently probed memory for the presented images in a surprise test. Our main prediction was that memory for images seen under continuous motion would be better than for images seen discontinuously, owing to temporal association mechanisms that are constrained by token assignment based on motion kinematics.

Experiment 1: Integrating Noisy Inputs on the Basis of Spatiotemporal Continuity

Real-world experience with objects is often noisy and impoverished. To create such conditions in the laboratory, in Experiment 1a we randomly assorted 50% of the pixels in the images shown during encoding. Crucially, this image noise was inserted on both appearances of a given image in a trial, but assorted independently each time. This allowed us to investigate how observers integrate information from noisy experiences to build robust memories. In Experiment 1b, we replicated effects with noise injected at test, but not during encoding. The critical manipulation, as already described, was whether the images were presented continuously or discontinuously. We expected better recognition memory for images seen in continuous motion, that is, in a single apparent motion stream.

Method

Participants. For each of the experiments reported our goal was to test and analyze results from approximately 20 participants. We sought 20 because a previously published exploratory study confirmed reliable effects and an effect size of $\eta_p^2 = 0.18$, using a sample size close to 20 (Schurgin, Reagh, Yassa, & Flombaum, 2013). That study reported an experiment identical to the current Experiment 1, except that it included only noiseless presentations and testing of images. A power analysis on the results of that experiment demonstrated that 18 participants would be sufficient, with a power of 0.95 and a 0.05 significance level.

We expected that engagement with the task would vary by subject and over the course of the several semesters during which experiments would be run. We therefore adopted a uniform exclusion criterion. Data from an individual subject was excluded from analysis if old-image classification accuracy (i.e., number correctly identified as old divided by number of old items tested) was more than two standard deviations below the group mean (with the subject included); or if more than 50% of responses were of a single response category (equal numbers of old, similar, and new images were presented during test).

Separate groups of 20 Johns Hopkins University undergraduates participated in Experiments 1a and 1b. The results from four participants were excluded in Experiment 1a. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus. Experiments took place in a dimly lit sound-attenuated room. Stimuli were presented on a Macintosh iMac computer with a refresh rate of 60 Hz. The viewing distance was 60 cm so that the display subtended 39.43° × 24.76° of visual angle.

Stimuli and procedure. Stimuli were generated using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented within the full display frame of 39.43° × 24.76°. Participants completed a visual object recognition memory task that included two stages. During the first phase (incidental encoding), participants were shown 384 color images of real-world objects on a computer screen with the cover task of indicating whether an item onscreen was an "indoor" or "outdoor" item. They indicated their responses using the computer keyboard.

Items in the task were shown using a common apparent motion paradigm, wherein an item could be seen as part of a single apparent motion stream or as parts of two different apparent motion streams (see Chun & Cavanagh, 1997 and Yi et al., 2008, upon which these methods were closely modeled). Each trial included four 200 ms frames with two images in each frame (see Figure 1). In the first frame, the two images consisted of scrambled object parts, and the images were positioned in the periphery to bias the perceived motion directions in subsequent frames. In the second frame, the images approached fixation, and one of them turned into a real-world object. In the third frame, the images passed the central region, and one of them turned into the same real-world object that appeared in the previous frame. Finally, in the fourth frame, the images moved to peripheral positions. Thus, in all trials, participants saw a single real-world object repeated twice, but either along the same stream (continuous motion) or between streams (discontinuous motion). Participants were told that they would see two images of the same object during each trial. We also instructed participants to do their best to maintain fixation in the center of the display, a strategy which we suggested would maximize performance. But we did not enforce fixation.

Experiment 1a manipulated the presentation of the real-world object in the second and third frame so that 50% of the pixels were randomly scrambled.

To enhance the perception of two distinct apparent motion streams, 20% of the trials contained only mask-transients in in all four frames. Participants were instructed to press the space bar if a trial contained no image of an object. Trails were self-paced and began when the participant pressed space to continue.

During each trial of the second phase (surprise retrieval), participants viewed a single image presented on the screen. The images were those presented in the previous task (old images), objects similar but not identical to ones in the previous task (similar images), and completely new objects that were not in the previous task (new images). For each image, participants were instructed to indicate whether it was "old," "similar," or "new." Participants were instructed to classify an image as old only if it was identical to the image they saw previously. They were also told that similar images would be categorically similar to an image they saw previously, but something may have changed. During the instructions for the surprise test, participants were shown examples (not from the exposure set) of old, similar, and new images to make these instructions clear. Images were shown for 3000 ms

each, with a 500 ms interstimulus interval. Participants could only make a response when an image was onscreen, and only the first response made was recorded. Overall, participants did not respond on 3.6% of trials in Experiment 1a and 4.7% in Experiment 1b.

In Experiment 1b, the stimuli and procedure were the same as in Experiment 1a with the following exceptions: At encoding, participants viewed a total of 360 color images of objects without noise. At retrieval, participants classified images of objects, the pixels of which were randomly scrambled by 25%, 50%, or 75%. The three levels of noise were equally distributed across old, similar, and new images.

Data analysis. We did not include responses that were faster than 200 ms, which accounted for a total of 0.66% of trials in Experiment 1a and 1.58% of trials in Experiment 1b.

Results

Cover task performance during encoding. During incidental encoding participants were asked to classify the image that appeared in each trial as either 'Indoors' or 'Outdoors.' To determine whether motion kinematics affected responses on this cover task, we computed the probability of a given classification as a function of motion type. There was no significant difference for continuous (indoor classifications, $M = 61.6\%$) compared with discontinuous motion (indoor classifications, $M = 60.4\%$), $t(383) = 0.25$, $p = .8$.

Recognition from memory during test. To de-confound discrimination ability from potential response biases, we analyzed recognition performance using a signal detection measure d_a (Green & Swets, 1966; Loiotile & Courtney, 2015). Each trial of the surprise test included either one of the objects shown during encoding, a novel object that was not shown previously, or a similar object: an object in the same category as and bearing some resemblance to one of the objects actually shown during encoding. Participants reported whether they thought an object was 'old,' 'new,' or 'similar,' that is, previously seen, unseen, or resembling something seen, respectively. With d_a we could compute both a Novel-Old discrimination index and a Similar-Old discrimination index.

We observed significantly better Novel-Old discrimination for objects encountered along spatiotemporally continuous ($d_a = 1.25$) compared with discontinuous ($d_a = 0.98$) paths, $t(15) = 3.00$, $p < .01$, Cohen's $D = 0.51$ (see Figure 2). A similar trend was observed for Similar-Old discriminations, although this effect did not reach significance. (Continuous $d_a = 0.27$; discontinuous $d_a = 0.17$, $t(15) = 1.65$, $p = .12$, Cohen's $D = 0.37$).

Memory strength analysis. To further investigate the extent to which exposure through spatiotemporal continuity supports the acquisition of more robust memory, we performed a memory strength analysis of the kind often used in research on long-term memory (Wickens, 2002, Ch. 3; Wixted, 2007). The analysis assumes that recognition decisions are based on the strength of a memory signal in relation to a decision criterion (Egan, 1958; Ratcliff, Sheu, & Gronlund, 1992). We conceived of the analysis as follows: each time an object image was encountered during encoding it added strength to a memory trace for that object. Was greater strength added when the second encounter followed the first under continuous motion (implying a single object token)?

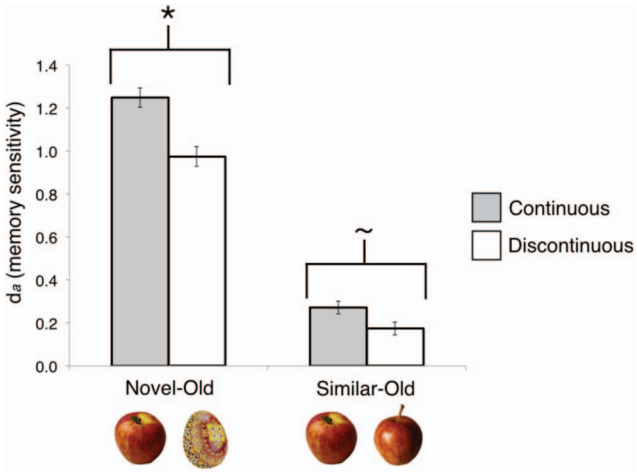

Figure 2. Results of Experiment 1a ($n = 16$). For both Novel-Old and Similar-Old comparisons, participants' discrimination performance was better for objects encountered along continuous as opposed to discontinuous paths. * designates $p < .01$, ~ designates $p = .12$. Error bars represent within-subject error (to remove between-subjects variability, see Cousineau, 2005). See the online article for the color version of this figure.

We could answer the question by modeling the strength of memory traces fit to classifications made during the surprise test.

The model assumes that in a continuous space of memory strength, test stimuli would elicit normal distributions centered around different means (μ) with unequal variance (Wixted, 2007). The distribution of responses elicited by New images was specified to be centered at 0 (i.e., on average, failing to elicit a memory signal since those images had not been seen previously) with a standard deviation (*SD*) of 1. For Old images, we created two distributions coded by motion continuity (continuous vs. discontinuous). These distributions had their means as free parameters, and a shared *SD*, also a free parameter. The distribution of memory strengths elicited by similar images also included a mean and *SD* as free parameters. To classify responses in the model, we specified free parameters for two ordered decision criteria (λ). Within a given distribution any sample above λ_1 would be classified as "Old," samples between λ_1 and λ_2 would be classified as "Similar," and samples less than λ_2 would be classified as "New." In total, this gave the model seven parameters to fit the data.

For simplicity of reporting the present model used averaged data across all participants as its input, although we also employed a hierarchical version to account for individual data. (Results were virtually identical with those of the aggregate model). The model simulated the averaged data for 10,000 iterations across two chains.

The results are shown in Table 1. Figure 3 plots the results graphically. The key effect is that the memory traces evoked by continuously presented items were stronger than were the memory traces evoked by discontinuously presented items. This difference was significant ($p < .01$); over the 10,000 simulations, the 99% confidence interval for the difference between μ (Continuous) and μ (Discontinuous) was 0.44 to 0.12 (thus larger than 0). This analysis adds to reported sensitivity comparisons by showing how differences in memory strength between continuously and discon-

Table 1
Parameter Outputs for Model of Experiment 1a

Parameter	Mean	Standard deviation	97.5% CI
μ (Continuous)	1.50	.06	1.61–1.38
μ (Discontinuous)	1.21	.05	1.32–1.11
SD (Cont. & Disc.)	1.50	.07	1.36–1.64
μ (Similar)	1.10	.04	1.17–1.03
SD (Similar)	1.13	.05	1.04–1.23
λ_1	1.61	.04	1.63–1.54
λ_2	.55	.02	.60–.51

Note. Outputs of parameter values for the memory strength model of Experiment 1a. Old items were shown either continuously or discontinuously, and were thus fit by distributions with different means. The main result is that the fitted μ for continuously shown items were larger than for discontinuously shown ones, implying stronger memory traces on average.

tinuously shown memoranda could have produced the particular patterns of response observed.

Experiment 1b

Experiment 1b was designed to replicate the main results of Experiment 1a, with a small methodological modification. In this case, images were presented without noise during encoding. But at test, images were embedded in either 25%, 50%, or 75% noise. In Experiment 1a, therefore, exposure to images was impoverished and the opportunity to recognize those images took place under relatively better conditions. In Experiment 1b, this was reversed: exposure conditions were better than test conditions, which varied from slightly noisy to very noisy.

Recognition From Memory During Test

Experiment 1b replicated the effects observed in 1a. A between-subjects ANOVA revealed a main effect of motion continuity on Novel-Old discrimination, $F(1, 19) = 19.12$, $p < .01$, $\eta_p^2 = 0.50$, and also a main effect of noise level, $F(2, 38) = 26.23$, $p < .01$, $\eta_p^2 = 0.58$, with no interaction between the two $F(2, 38) = 0.19$, $p = .83$, $\eta_p^2 = 0.01$. As noise level increased, performance decreased. Across all three noise levels there was a benefit for motion continuity (see Figure 4). Planned comparisons contrasting continuous and discontinuous exposure were significant at all three noise levels (all $p < .05$). As in Experiment 1a, effects on Similar-Old discriminations were in the same direction, but failed to reach significance.

Memory Strength Analysis

We again applied a memory strength model to analyze the results of Experiment 1b. The same model was applied here as in Experiment 1a, but it was applied three times, once to the data from each noise condition at test. Table 2 shows μ parameter values for continuous and discontinuous images as a function of noise level, along with 95% confidence intervals for their differences. Again, continuous motion elicited stronger memory traces than discontinuous (all $p < .05$).

Experiment 1b also provided an interesting test of concept for the memory strength model. Additional noise at test should cause an image to evoke a weaker memory trace (when it was seen previously). Indeed, mean μ values as a function of noise level declined, particularly for items with 75% noise injected compared with 50%. This effect is less a prediction of the model, and more a useful demonstration that the model generally captures effects that should impact the strength of an evoked memory signal.

Together, the novel effects reported in Experiment 1 support the hypothesis that token identity in the short-term is exploited to support long-term object memory. These experiments do not identify the exact mechanisms by which token identity is tracked and then imposes its constraints. We discuss potential mechanisms, including the role of attention, in the General Discussion under the heading, 'Potential Mechanisms and the Role of Attention.' Ex-

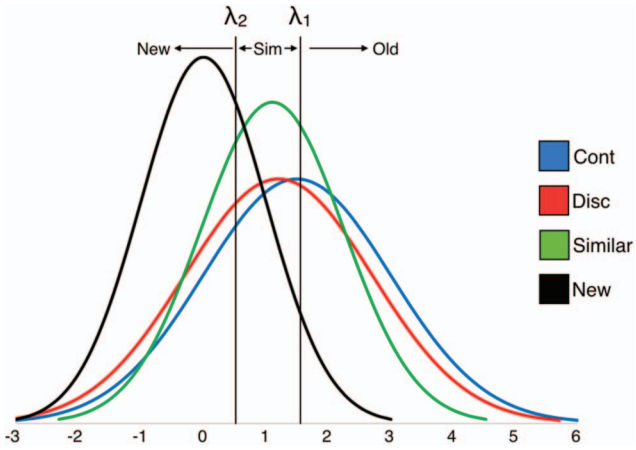

Figure 3. Fitted model distributions. Under the assumption that new items evoke virtually no memory trace, those (shown in black) were assigned a μ of 0 and a *SD* of 1. Distributions for continuously presented items (blue [dark gray]) were fit, along with distributions for discontinuous items (red [gray]) and similar items (green [light gray]). Continuous items elicited stronger memory traces, on average. See the online article for the color version of this figure.

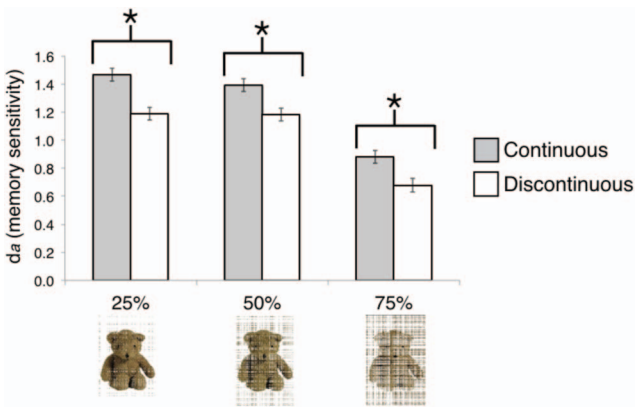

Figure 4. Results of Experiment 1b ($n = 20$). At test, participants judged whether an image was old, similar, or new relative to objects encountered during encoding. Each image was embedded in either 25, 50, or 75% noise. Across all three noise levels discrimination of Old objects was better when the object was encountered along a continuous (compared with a discontinuous) path. * designates $p < .05$. Error bars represent within-subject error. See the online article for the color version of this figure.

Table 2
Subset of Outputs for Model of Experiment 1b

Parameter	Noise level 25%	Noise level 50%	Noise level 75%
μ (Continuous)	1.57	1.55	1.15
μ (Discontinuous)	1.25	1.33	.93
95% confidence μ (Cont-Disc)	.51 to .13	.41 to .028	.45 to .012

Note. μ values as a function of noise level for continuously and discontinuously presented images, along with 95% confidence intervals for their differences. Continuously presented items elicited stronger memory signals when represented at test across all noise conditions. [Full parameter outputs for this and all reported experiments can be obtained at http://www.jhuvisualthinkinglab.com/demos-schurgin-and-flombaum-stcontinuity].

periment 2b will additionally address the potential role of attention.

Experiment 2: Replication With a Forced Choice Test

Experiments 1a and 1b investigated an impact of motion kinematics during encoding, and for this reason, we utilized an old/similar/new procedure that has been used previously in studies of long-term memory investigating encoding effects (Bakker et al., 2012; Kim & Yassa, 2013; Rentz et al., 2013; Stark, Yassa, Lacy, & Stark, 2013). Because the effects reported are novel, and involve a kind of manipulation not typical in research on long-term memory (i.e., perceived motion), we were motivated to replicate the effects with a forced choice measure of long-term memory. According to some theories, old/similar/new judgments with a single test image index different processes than forced choices between two or more images. Thus, Experiment 2a was designed to replicate the effects in Experiment 1a while using a forced choice procedure to demonstrate the reproducibility and pervasiveness of these effects. Experiment 2b used the methods of 2a to serve as a control investigating an alternative attention-driven account of the results.

Method

Participants. A new group of 21 Johns Hopkins University undergraduates participated in Experiment 2a and another group of 21 participated in Experiment 2b. The results from two participants were excluded from Experiment 2a following the exclusion criteria from Experiment 1. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, apparatus, and procedure. All methods were identical to those in Experiment 1a, with the following exceptions: At incidental encoding, participants were shown 368 images of real-world objects. At retrieval, participants performed a two alternative forced choice (2AFC) task. In the center of the screen were two objects, only one previously shown during incidental encoding. Participants had to indicate which of the two they had encountered during encoding. Half of these trials were Novel-Old comparisons, in which a previously shown object was paired with a new, categorically distinct object. The other half of these trials were Similar-Old comparisons, in which a previously shown object was paired with a similar-looking object from the same category. Each pair of images was present on the screen until the participant made a response.

Data analysis. We did not include responses that were faster than 200 ms, a total of 3.6% of trials.

Results

Cover task performance during encoding. As in Experiment 1, we found no significant difference in the proportion of outdoor/indoor image judgments during encoding as a function of motion continuity, $t(367) = 1.38$, $p = .17$.

Recognition from memory during test. We observed a main effect of motion continuity for Novel-Old comparisons, $t(18) = 2.24$, $p = .038$, Cohen's $D = 0.40$. Performance was better when objects were encountered along a spatiotemporally continuous ($M = 76.7\%$) versus discontinuous path ($M = 72.9\%$).

For Similar-Old comparisons, we observed a similar trend, $t(18) = 1.9$, $p = .07$. Performance was better when objects were encountered along a spatiotemporally continuous ($M = 59.2\%$) versus discontinuous path ($M = 55.5\%$), Cohen's $D = 0.57$. Figure 5 plots these results graphically.

Experiment 2b

A critical feature of our theorizing around the three experiments presented so far is that continuous motion—and consequent token perception—enhance memory performance by supporting the *integration* of independent encounters with the images shown. A salient set of questions about the mechanisms underlying such a process concerns the role of attention: might the mechanism of integration be an attentional bias induced by the initial appearance of an image in one of the streams? We will discuss this possibility in the General Discussion, because it seems likely and it is fundamentally consistent with our view that Core Knowledge inter-

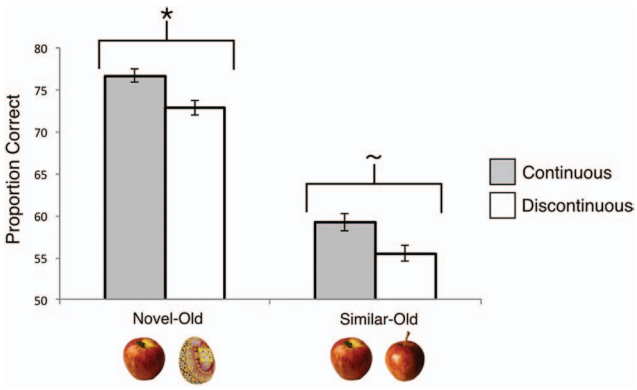

Figure 5. Results of Experiment 2a ($n = 19$). For both Novel-Old and Similar-Old comparisons, participants' discrimination performance was better for objects encountered along continuous as opposed to discontinuous paths. * designates $p < .01$, ~ designates a trending difference ($p = .07$). Error bars represent within-subject error. See the online article for the color version of this figure.

acts with psychological mechanisms broadly, including attention. Via Experiment 2b, however, we sought to control for an alternative account by which attention could conceivably act to produce the effects without also supporting integration over time.

Specifically, we were concerned that continuous motion merely biases spatial attention, causing the second image to be seen—and thus better remembered—without promoting integration of the encounters. Suppose that the first encounter with an image is often barely perceived or processed, because its location is unpredictable and its presence fleeting. Then that encounter may fail, on its own, to produce any substantive memory trace. But if attention is attracted by the first object, it could track the stream in which that object appears, which means it could more effectively capture the second appearance within continuous streams, while missing (or processing less so) any second appearance that falls in the discontinuous stream.[1] In other words, perhaps the benefit we observed is simply one of perceiving and processing a single image during its second appearance, not one of integrating two encounters, an effect having nothing to do with integration over token identity.

To control for this possibility, in Experiment 2b we placed two distinct objects in each trial. We manipulated motion continuity as we had previously, and then in the second phase of the experiment we tested memory for the second images shown during encoding (using the 2AFC methods of Experiment 2a). The alternative attention account just outlined predicts that memory for the second images within continuous streams would be better than for the second images within discontinuous streams. But in this case the effect would not be attributable to integration across encounters, because the relevant encounters always involved distinct images. This would be a problem because the modulation in memory performance would have nothing to do with integration of object knowledge from multiple observations. In contrast, a null result would suggest that continuity does not merely promote memory for images that follow other images in a continuous stream, and therefore, that the effects already observed arise in virtue of integration per se.

Because our point of view predicts a null result for this experiment, while the alternative predicts a significant effect, we used methods here that we hoped would produce good performance overall, maximizing the possibility of observing an effect and reducing the possibility of observing floor performance. Specifically, we used the 2AFC test of Experiment 2a, a test that generally provides for better performance than old/similar/new classification. For the same reason, we included only noiseless images at both encoding and test. We also restricted testing to old versus new images because similar discriminations in the previous three experiments only produced trending effects and generally low performance. Finally, the first image in each trial of incidental encoding was always the same picture of an apple. This was done to reduce the chances that the image would be integrated with the second image in each trial (which would produce poor memory representations), and to make it highly predictive with respect to the appearance of the second image. To the extent that the first image is a cue toward tracking continuity, we sought to make it as reliable and recognizable as possible. Figure 6 schematizes the methods of this experiment.

Method

Stimuli, apparatus, and procedure. The experiment was identical to 2a, with the following exceptions. For both continuous and discontinuous trials, the first image of a real-world object was always the same: an apple. Participants were instructed to judge whether the object that followed the apple was an "indoor" or "outdoor" object. No noise was present in the images during encoding. At test, the 2AFC task exclusively included Novel-Old comparisons.

This experiment was preregistered with aspredicted.org, available permanently at https://AsPredicted.org/85qv5.pdf. This experiment was conducted following a round of reviews, and thus after all of the other reported experiments. The lab recently adopted preregistration as standard practice, though this was after the other experiments had already been completed.

Data analysis. We did not include responses that were faster than 200 ms, a total of 0.06% of trials.

Results

Cover task performance during encoding. As in Experiments 1 and 2a, we found no significant difference in the proportion of outdoor/indoor image judgments during encoding as a function of motion continuity, $t(367) = 0.94$, $p = .35$.

Recognition from memory during test. We failed to observe a main effect of motion continuity for Novel-Old comparisons, $t(20) = 0.82$, $p = .92$, Cohen's $D = 0.03$. Performance was indistinguishable statistically regardless of whether an image appeared in the second position during encoding along a spatiotemporally continuous ($M = 86.1\%$) versus discontinuous path ($M = 85.9\%$). Figure 6b shows these results graphically.

Learning to ignore? What if in the control experiment, the repeated and recognizable apple caused participants to learn to ignore motion continuity, and this is why performance in the two conditions became similar (in contrast with Experiment 2a)? To rule out the possibility we analyzed performance over the course of the experiment in four blocks. Learning to ignore would predict a continuity effect early on in the experiment, and no effect only later. We found overall that performance slightly decreased as the experiment progressed: 87.63% in block 1, 87.50% in block 2, 85.38% in block 3, and 83.34% in block 4. The difference between block 1 and 4 was significant, $t(20) = 2.50$, $p = .02$. But we did not find a continuity effect in any individual block (all ps > 0.64).

Experiment 3: Integrating Inputs With Different Orientations

Changes to an observer's viewpoint—which result in changes to an object's orientation relative to the observer—exert radical changes on the inputs to the visual system from a given object. Consider the two images of a horse shown in Figure 1: the two images are extremely different in the particular stimulations that they produce on the retina. Yet we easily recognize the horse in the two images as the same—an

[1] We thank an anonymous reviewer for bringing the possibility of this mechanism to our attention and for suggesting the outline of Experiment 2b as a test. We have paraphrased his or her comments in the preceding sentence and in the following paragraph.

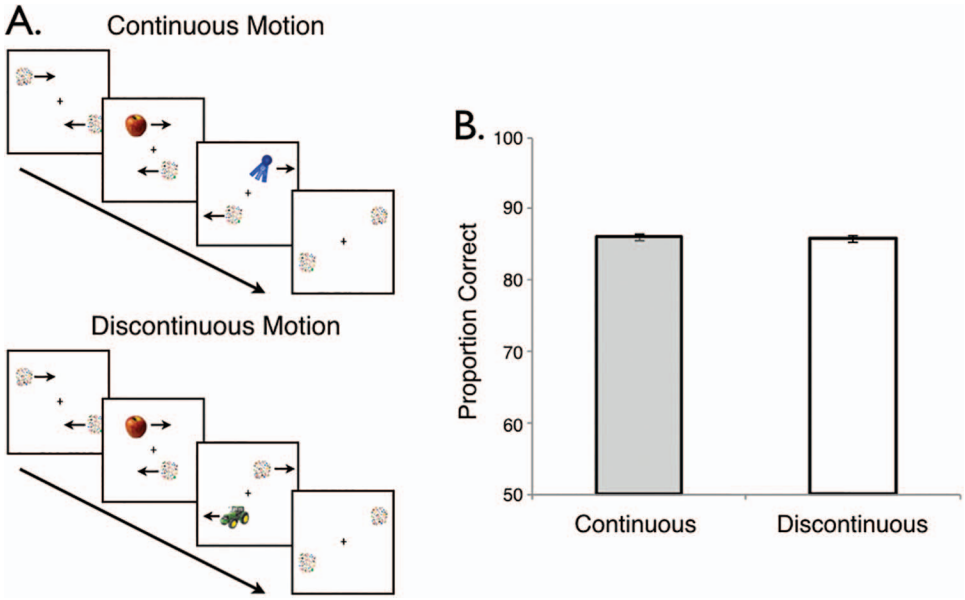

Figure 6. Methods and Results of Experiment 2b (n = 21). For Novel-Old comparisons, participants' discrimination performance was indistinguishable for images following an apple along continuous and discontinuous paths. Error bars represent within-subject error. These results are inconsistent with an account in which continuity merely promotes memory for images in the second position of a continuous path. See the online article for the color version of this figure.

ability often called viewpoint invariance, or tolerance. Experiment 3 was designed to directly test whether viewpoint invariance is supported by spatiotemporal continuity. Objects moving through space will naturally appear in changing orientations relative to an observer. Knowing that an object is nonetheless the same token can support integration in the service of invariance.

Experiment 3 was identical to Experiment 1, with the following exceptions. Each presentation of an object during encoding comprised two images of the objects at two different orientations (without image noise injected). At test, 'Old' objects were shown at a third, different orientation. To simplify the instructions to participants, this experiment included only 'New' foils, thus requiring that participants judge each image shown as either 'Old' or 'New.'

Method

Participants. A new group of 30 Johns Hopkins University undergraduates participated in Experiment 3. We initially recruited 20 participants, but found that our exclusion criteria demanded the removal of six. The effects with 14 were significant, but we decided to test an additional 10 participants to verify the findings considering so many exclusions. This resulted in one more participant meeting our exclusion criteria, for a total of seven participants removed and 30 tested total. Below we report results for the 23 included participants as well as the full set of 30, in the interest of transparency. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, apparatus, and procedure. In Experiment 3, the stimuli and procedure were the same as in Experiment 1 with the following exceptions: At encoding, participants judged whether the object was more "square" or "round." This change was made because the stimulus set used included an overwhelming number of objects that would have been rated as "indoor" objects. Participants judged a total of 316 color images of objects (taken from Geusebroek, Burghouts, & Smeulders, 2005). At retrieval, participants saw images that were either a new object or an old object from the incidental encoding phase but shown at a completely new orientation. They were asked to classify the images at test as "Old" or "New."

Data analysis. At retrieval, responses were only recorded if the image was still present on the screen. Overall, participants did not respond on 1.5% of trials. Additionally, we did not include responses that were faster than 200 ms, which resulted in an additional 0.3% of trials being excluded from the reported analyses.

Results

Cover task performance during encoding. As in Experiments 1–2, we found no significant difference in the proportion of outdoor/indoor image judgments during encoding as a function of motion continuity, $t(157) = 1.02$, $p = .31$.

Recognition from memory during test. Requiring only an Old or New judgment meant that we could compute d' as a measure of sensitivity. As in Experiment 1, performance was better when images appeared along continuous as opposed to discontinuous paths ($d' = 1.35$ vs. $d' = 1.22$, $t(22) = 2.67$, $p = .01$, Cohen's D = 0.23). When including all 30 participants, results were still significant ($d' = 1.24$ vs. $d' = 1.14$, $t(29) = 2.34$, $p = .026$, Cohen's $D = 0.19$). Thus, even when recognizing an object at a third, never-before-seen viewpoint, spatiotemporally continuous motion supported better recognition performance (Figure 7).

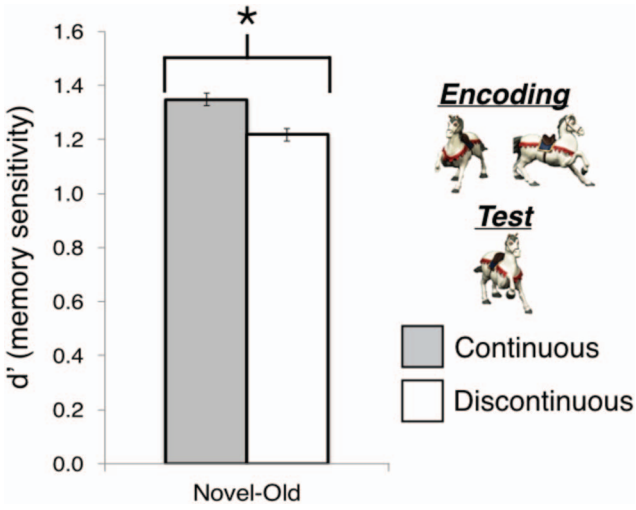

Figure 7. Results of Experiment 3 ($n = 23$). Participants' memory sensitivity was better for objects encountered along continuous as opposed to discontinuous paths. * designates $p < .05$. Error bars represent within-subject error. See the online article for the color version of this figure.

Memory strength analysis. For this experiment, we again performed a model-based memory strength analysis. Recall that during presentation each exposure to an object was at a different orientation. The theory in this case is therefore that continuity (compared with discontinuity) would breed integration across these orientation differences, thus producing stronger memory signals when the objects were seen during test at a third orientation. This experiment did not include similar foils or similar judgments. Thus, it could be modeled with fewer free parameters, specifically the μ and *SD* for continuous and discontinuous old object distributions, and a single λ criterion. Average μ estimates for continuously perceived objects (1.84) were higher than for discontinuously perceived objects (1.57). The 95% confidence interval for the differences between the condition estimates was .66 to .01.

Experiment 4: Integrating Noisy Inputs Through Occlusion and Disocclusion

Research has shown that infants and human adults track object persistence via multiple cues, and that they track object persistence despite occlusion, the complete disappearance of an object caused by another object in-between it and the viewer (Burke, 1952; Spelke et al., 1995; Xu & Carey, 1996; Yi et al., 2008). In Experiment 4 we sought to conceptually replicate the results of Experiment 1 while employing a different manipulation of motion continuity.

Again, the experiment began with an incidental encoding phase, and again each trial included a single image that appeared twice. Here, however, the appearances of the images followed two episodes of complete occlusion. In the continuous motion trials, the images appeared from behind the *same* occluder (a pillar drawn into the image). While in the discontinuous motion trials the images appeared from behind two *different* occluders, each presented on opposite sides of the display. And crucially, the image never fully traversed the space between the occluders. In both conditions, motion was fast in the periphery and slower at fixation so that the majority of the presentation of each image was directly near fixation regardless of the motion continuity condition. Figure 8 schematically illustrates these presentations.

Method

Participants. A new group of 22 Johns Hopkins University undergraduates participated in Experiment 4. The results from two participants were excluded. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, apparatus, and procedure. In Experiment 4, the procedure was the same as in Experiment 1, except as follows. During each trial of incidental encoding an image appeared twice under dynamic occlusion. Participants judged the object as either 'indoors' or 'outdoors' (as in Experiment 1). The encoding display contained a vertical column on each side of fixation (each occupying a space of 6.4° and starting 10.6° from fixation on either side). An image embedded in 50% noise appeared from behind one of the two columns, moved to fixation, reversed motion direction to return to its origin column, and it then disappeared by naturalistic occlusion behind that same column. Critically, the same image appeared a second time embedded again in (independent) 50% noise. But the second appearance could be either from behind the same (continuous) column as the original appearance, or the one on the opposite side of the display (discontinuous).

Subjects were instructed to fixate the central cross throughout the experiment, although eye movements were not monitored. The images always exited an occluder at a speed of 38.72°/s During each approach to fixation, the speed (S_i) on each frame of subsequent motion (i) was equal to:

$$S_i = S_{i-1} - \left(0.64 - \left(\frac{0.02i}{4}\right)^2\right)$$

This produced a speed at fixation of 3.78°/s (30 frames of motion). Returning to its origin the object moved through the same trajectory in reverse. Thus, movement was very fast in the periphery and slower at fixation so that the majority of the presentation of each image was directly near fixation.

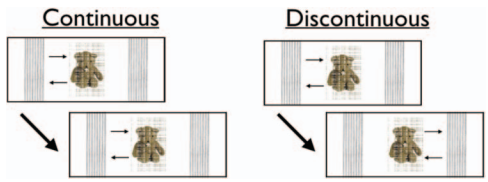

Figure 8. Procedure of the incidental encoding task used in Experiment 4. In each trial, an image of an object (embedded in noise) appeared from behind an occluder, moved toward the center of the display and then retreated to its origin, where it became occluded. This full trajectory lasted for one second, and it was repeated twice in a trial. The manipulation was whether the second appearance was from behind the same occluder as the first (continuous motion) or from the occluder on the opposite side of the display (discontinuous motion). See the online article for the color version of this figure.

Throughout the incidental encoding task participants viewed a total of 320 unique color images (i.e., there were 320 trials of incidental encoding). During the retrieval test, participants were asked to classify the objects as "Old" or "New." It was explained that they should only classify old images as "Old," and that new or similar looking objects should be classified as "New." The proportion of trials that contained old, similar and new images were exactly the same (160 images per image type). Compared with Experiments 1–3, we changed the response type (removing the similar response) because performance with that response type tended to be poor and to make the instructions more straightforward for participants. Each image was present on the screen until the participant made a response.

Data analysis. We excluded from analysis responses that were faster than 200 ms, a total of 1.3% of all trials.

Results

Cover task performance during encoding. As in Experiments 1–3, we found no significant difference in the proportion of outdoor/indoor image judgments during encoding as a function of motion continuity, $t(319) = 0.80$, $p = .43$.

Recognition from memory during test. Participants demonstrated extremely high performance for Old images overall, $d' = 2.44$. This advantage compared with the previous experiments was likely caused by the greatly increased encoding time in Experiment 4, with each image remaining present for a total of 2000 ms (compared with 400 ms in Experiments 1–3).

As in the previous three experiments, we found a significant performance benefit for Old images that had been encoded under continuous motion ((d') = 2.48) compared with objects encountered under discontinuous motion ((d') = 2.41), $t(19) = 2.20$, $p = .04$ Cohen's $D = 0.13$.

Memory strength analysis. We also applied a memory strength model to this experiment. estimated μ values were much higher than before—again, probably because of the greatly increased exposure time and concomitant performance. But μ estimates for continuously presented images were on average larger than for discontinuously presented ones (mean of the estimates 2.37 v 2.30, respectively). And the 95% confidence interval for the mean estimate of the parameter differences slightly fell below 0, having a range between .22 and −.031. The smaller effects in this experiment make sense given the increased exposure times and higher performance, and they are also consistent with previous work by Yi and colleagues (2008) on the effects of motion continuity, wherein the same apparent motion manipulation used in the previous experiments had larger and more reliable effects than the same occlusion manipulation used in this experiment.

Discussion

The problem of object recognition is often characterized as one of invariance, storing memories that can be used to recognize an object despite drastic changes at the level of basic inputs. Any given object can induce a wide array of retinal stimulations because of factors as basic as lighting conditions, viewer motion, and rotation. To meet this challenge one strategy involves temporal association—integrating noisy experiences during one encounter to build representations that tolerate variability at later encounters. Our results demonstrate that temporal association is a component of human object memory, but critically, one that is supported and constrained by core principles of spatiotemporal object persistence. Tracking an object at the level of a 'token' allows an observer to leverage changes in appearance over the short-term to produce appropriate discrimination in the long-term.

Two important features of our experiments are the use of both low-level image noise (Experiments 1 and 4) and an orientation manipulation (Experiment 3), as well as the use of two rather different manipulations of object continuity (occlusion and apparent motion). These different manipulations produced results that varied in magnitude, but importantly, the similarity in the kinds of effects obtained make many lower-level factors unlikely as explanations. Specifically, we obtained effects with very short exposures (400 ms) and with relatively longer exposures (2 s); we obtained the effects when the relevant stimuli could be construed as competing with other stimuli in the display for encoding (Experiments 1–3), and when they appeared alone in the display (Experiment 4); we obtained effects for stimuli that could be easily fixated, regardless of continuity condition (Experiment 4), as well as stimuli that might have been viewed often outside of the fovea (Experiments 1–3). This convergence suggests that the effects are not exclusive to taxing encoding settings or the opposite, to more comfortable ones. Nor are they exclusive to instances in which eye movements cannot be made quickly enough to directly pursuit a moving object—in Experiments 1–3, each image appeared for only 200 ms, followed immediately by the next image while a saccade takes at least 200 ms—nor to settings in which eye movements could easily adjust following an incorrectly anticipated image appearance—in Experiment 4, each image remained present and smoothly moving for one second, and it was always the only image in the display. We also obtained the same effects using a variety of testing procedures and analyses, including alternative forced choice, old/similar/new judgments, signal detection measures, and memory strength modeling. Effects of spatiotemporal continuity on object encoding and subsequent memory appear, therefore, to reveal a general mechanism for associating inputs based on tracked token identity.

Potential Mechanisms and the Role of Attention

Future research will be required to characterize the exact nature of the mechanisms involved in the process of using token identity to constrain object memory. It seems to us relatively uncontroversial to believe that the relevant mechanisms operate without intention to encode and remember nor even the intention to track, in any explicit sense.

One promising candidate for a supporting mechanism is visuospatial attention. Attention is known to constrain learning and memory in a variety of specific cases (Chun & Turk-Browne, 2007; Kawachi & Gyoba, 2006; Scholl & Pylyshyn, 1999). But there are several different ways that attention could play a role in the experiments, each with different implications.

In the context of Experiment 2b, a control, we discussed one alternative account of our results that would appeal to attention. This involved an attentional bias toward continuous motion, but one that does not promote integration across encounters. Note that even that account has built into it a reliance on spatiotemporal continuity. And because Experiment 4 (dynamic occlusion) was so

different in kinematics and structure from the other experiments, a consistent alternative account would require several ad hoc adjustments to explain all our results. (e.g., Experiment 4 involved 2000 ms exposure durations compared with only 400 ms exposures in the other experiments, making it unlikely that a second object appearance could be completely missed, even given an attentional bias). More importantly, though, Experiment 2b dispatched this alternative account empirically, and so we turn now to another attentional account that seems both more plausible and includes a role for integration through motion.

An image appearing in one place could conceivably produce an expectation of an image in a second, trajectory-extrapolated place, directing attention automatically, or at least without volition necessary, to that second location. On this view, spatiotemporal expectation would be conceived as built into the operation of attention, producing a sort of filter by which association is naturally biased toward signals that are likely to have arrived from the same source. Crucially, the view is one that involves association, not merely more exposure. Viewing the role of attention in this way might suggest that token identity is not tracked explicitly or intelligently, but more implicitly and ballistically. However, the argument that the relevant expectations are hardwired, as opposed to representation-dependent, becomes less appealing considering the abundance of evidence that attention has been shown to have the right expectations about token identify in very different contexts (see the discussion called 'Core Knowledge Supports Cognition,' below). We are comfortable with either interpretation, to be sure, noting that the important contribution at this moment is that token identity, in practice, affects object memory.

An alternative to an attentional account would be that even with diffuse or unfocused attention, integration of signals is constrained after entering memory encoding. For example, whether a second image is related to a first via pattern completion or pattern separation in the medial temporal lobe (MTL) could be a 'switch' that token identity acts upon (Yassa & Stark, 2011). Indeed, these two computations in the MTL seem to bear an uncanny resemblance to the opposing challenges of tolerance and invariance often discussed in the context object recognition. Thus, it is surprising that long-term visual memory and object recognition are often treated as separate topics.

At present our results are insufficient to adjudicate between a more memory- or attention-based account of the influence of token identity. In lieu, we therefore emphasize that the results are not mutually exclusive, and that there are extant reasons to anticipate that both hold some truth. Importantly, both hypotheses suggest an interaction between the underlying mechanisms of perception and memory, topics which are often investigated in isolation.

Implications for "What" and "Where" in Object Recognition

Similarly, our results suggest that human object recognition depends on the coordination of 'what' and 'where' representations of objects. These representations are often ascribed to ventral and dorsal processing streams, respectively, with research on object recognition generally focused on the former. In our experiments, continuous motion supported the assimilation of information about object appearances, suggesting that investigating coordinated activity in these systems could lead to advances. Research on long-term visual memory rarely involves manipulations designed to engage token processing mechanisms. Most visual long-term memory studies display only static stimuli in fixed positions on a screen (Brady, Konkle, & Alvarez, 2011; Guerin, Robbins, Gilmore, & Schacter, 2012; Kim & Yassa, 2013; Stark et al., 2013). If memory mechanisms are optimized for, expect, and/or depend upon at least some object motion and related stimulus variability, then static objects may fail to fully engage typical encoding mechanisms. Integrating this knowledge into subsequent research could lead to advances in our understanding of long-term memory.

The Importance of Token Identity Beyond Perception

Our results also suggest that clues to object recognition may generally be found in research more typically framed in terms of attention and perception. For example, in the domain of object file research (using a paradigm known as object reviewing) some studies have investigated a small set of appearance transformations that occasionally *preclude* token ascription. It has also been found that under certain circumstances similarities in physical appearance engender token ascriptions, despite a lack of any obvious cues to spatiotemporal continuity (Hollingworth & Franconeri, 2009; Mitroff & Alvarez, 2007; Moore, Stephens, & Hein, 2010; Richard, Luck, & Hollingworth, 2008). Presumably any expectations about the physical appearances of objects that are strong enough to guide token perception will play an important role in scaffolding long-term recognition and categorization.

Implications for Machine Learning

These results suggest a remedy to the problem of unsupervised learning in object recognition. In machine learning, object recognition is often trained with supervision—an oracle explicitly conveys which exemplars in a training set belong to the same categories or are instances of the same object (Andreopoulos & Tsotsos, 2013). Thus, machine learning algorithms demand prior knowledge (and often, a great deal of training). What could play the role of supervision in human object recognition, especially in very young children? Tracking of token identity can be thought of as supplying a sequential set of test stimuli with feedback to a learning program. If one knows that an object at time point B is the same token as the one at time point A, then one can ask whether the image observed at B is consistent with the representation observed at A. If it is, then it should reinforce the representational structure acquired at A. If it is not, it should lead to an adjustment of the representation and a new prediction about what should be encountered at time point C.

This is largely the point that has been made in research on how eye movements can support learning with a temporal association rule: it is a good bet that the intended target of a saccade and its final terminus are instances of the same token (Cox et al., 2005; Isik et al., 2012; Li & DiCarlo, 2008). Perhaps learning through saccades is a particular case of a more general strategy of learning about appearance by tracking token identity. We are not aware of any research in young children or infants that has investigated temporal association through saccades in the interest of object memory. But a great deal of research shows that children and infants track token identity over time, largely by relying on spa-

tiotemporal constraints (Baillargeon, Spelke, & Wasserman, 1985; Spelke et al., 1995; Stahl & Feigenson, 2015; Xu & Carey, 1996). Thus, learning about object appearance through token assignment is a viable mechanism for early learning in the service of object recognition.

Core Knowledge Supports Cognition

Finally, our results demonstrate that mechanisms underlying 'Core Knowledge' may be a critical component of object recognition and long-term memory. This is consistent with the broader literature, which has long established that Core Knowledge constrains memory, learning and perception across the life span (e.g., Cheries, Mitroff, Wynn, & Scholl, 2008; Stahl & Feigenson, 2015; Van Marle & Scholl, 2003). Here we have focused on knowledge of spatiotemporal persistence (at times, called 'continuity') in the context of object motion and occlusion. Core Knowledge also includes an understanding of object cohesion (Cheries et al., 2008; Noles, Scholl, & Mitroff, 2005; Xu & Carey, 1996), the specific dynamics of occlusion (Scholl & Pylyshyn, 1999) expectations about balance (Baillargeon & Hanko-Summers, 1990), and possibly even expectations about entropy (Newman, Keil, Kuhlmeier, & Wynn, 2010), among other things. In many of these instances infants and adults not only possess the relevant knowledge, but the corresponding expectations also appear to govern reasoning and learning. For example, there is evidence that infants reason about the number of objects in a scene based on spatiotemporal expectations (Wilcox & Baillargeon, 1998a, 1998b; Xu & Carey, 1996). Violations of object cohesion have been shown to corrupt infants' expectations about more versus less (Cheries et al., 2008), while the same kind of cohesion violations have been shown to disrupt object file priming (Noles, Scholl, & Mitroff, 2005). Violations of naturalistic occlusion impair multiple object tracking in human adults (Scholl & Pylyshyn, 1999). And in adults, the tendency to classify otherwise similar physical events as involving occlusion or containment influences which features are more easily remembered in a working memory paradigm (Strickland & Scholl, 2015). Thus, Core Knowledge appears to be a set of constrains on cognition with a domain general reach.

Surprisingly though, long-term memory and object recognition have not been considered previously as domains influenced by Core Knowledge. A recent study by Stahl and Feigenson (2015) is perhaps the exception, where violations of expectations in infants were shown to influence associations between objects and features. Given those results and the ones presented here, the natural hypothesis that follows is that Core Knowledge in general plays an important role in long-term visual memory. Indeed, the motivation for our experiments was that expectations about persistence create an opportunity for appropriately constrained memory (i.e., memory that balances tolerance and discrimination); if one knows that an object is still the same, one can learn just how different that object can look form itself.

Similar opportunities may arise by applying expectations about other aspects of Core Knowledge. For example, one challenge in object recognition is to distinguish between objects that are distinct, but touching or otherwise contiguous in a two-dimensional image plane. Research on balance and support suggests that even young infants have expectations about when one object could conceivably support a contiguous one, showing surprise when an object does not topple over under certain relationships (Baillargeon & Hanko-Summers, 1990). Knowing something about when contiguous parts of an image are distinct objects should affect long-term representations of those objects. Broadly then, we hope that Core Knowledge is integrated into research on how visual long-term memories are acquired and encoded because that knowledge may supply solutions to many of the computational challenges facing object recognition.

Conclusion

Altogether, perhaps the main conclusion to draw is that tolerance in long-term object recognition arises from a kind of over-tolerance in shorter-term object recognition. Confidence about token identity based on kinematics lends flexibility to perceptual and working memory mechanisms so that changing inputs can become amalgamated.

References

Andreopoulos, A., & Tsotsos, J. K. (2013). 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding, 117,* 827–891. http://dx.doi.org/10.1016/j.cviu.2013.04.005

Anstis, S. M., & Mackay, D. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 290,* 153–168. http://dx.doi.org/10.1098/rstb.1980.0088

Baillargeon, R., & Hanko-Summers, S. (1990). Is the top object adequately supported by the bottom object? Young infants' understanding of support relations. *Cognitive Development, 5,* 29–53. http://dx.doi.org/10.1016/0885-2014(90)90011-H

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition, 20,* 191–208. http://dx.doi.org/10.1016/0010-0277(85)90008-3

Bakker, A., Krauss, G. L., Albert, M. S., Speck, C. L., Jones, L. R., Stark, C. E., . . . Gallagher, M. (2012). Reduction of hippocampal hyperactivity improves cognition in amnestic mild cognitive impairment. *Neuron, 74,* 467–474. http://dx.doi.org/10.1016/j.neuron.2012.03.023

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147. http://dx.doi.org/10.1037/0033-295X.94.2.115

Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. *The Journal of Neuroscience, 26,* 916–925. http://dx.doi.org/10.1523/JNEUROSCI.2353-05.2006

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11,* 4. http://dx.doi.org/10.1167/11.5.4

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science, 281,* 1185–1187. http://dx.doi.org/10.1126/science.281.5380.1185

Burke, L. (1952). On the tunnel effect. *The Quarterly Journal of Experimental Psychology, 4,* 121–138. http://dx.doi.org/10.1080/17470215208416611

Cheries, E. W., Mitroff, S. R., Wynn, K., & Scholl, B. J. (2008). Cohesion as a constraint on object persistence in infancy. *Developmental Science, 11,* 427–432. http://dx.doi.org/10.1111/j.1467-7687.2008.00687.x

Chun, M. M., & Cavanagh, P. (1997). Seeing two as one: Linking apparent motion and repetition blindness. *Psychological Science, 8,* 74–79. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00686.x

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology, 17,* 177–184. http://dx.doi.org/10.1016/j.conb.2007.03.005

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1,* 42–45. http://dx.doi.org/10.20982/tqmp.01.1.p042

Cox, D. D., & DiCarlo, J. J. (2008). Does learned shape selectivity in inferior temporal cortex automatically generalize across retinal position? *The Journal of Neuroscience, 28,* 10045–10055. http://dx.doi.org/10.1523/JNEUROSCI.2142-08.2008

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nature Neuroscience, 8,* 1145–1147. http://dx.doi.org/10.1038/nn1519

Dennett, D. C. (2006). Cognitive wheels: The frame problem of AI. In J. Bermudez & F. MacPherson (Eds.), *Philosophy of psychology: Contemporary readings* (pp. 433–454). New York, NY: Routledge.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11,* 333–341. http://dx.doi.org/10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73,* 415–434. http://dx.doi.org/10.1016/j.neuron.2012.01.010

Egan, J. P. (1958). *Recognition memory and the operating characteristic*. USAF Operational Applications Laboratory Technical Note, 58–51.

Flombaum, J. I., Kundey, S. M., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychological Science, 15,* 795–800. http://dx.doi.org/10.1111/j.0956-7976.2004.00758.x

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. http://dx.doi.org/10.1037/0096-1523.32.4.840

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. *The origins of object knowledge,* 135–164.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT press.

Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. (2005). The Amsterdam library of object images. *International Journal of Computer Vision, 61,* 103–112. http://dx.doi.org/10.1023/B:VISI.0000042993.50813.60

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York, NY: Wiley.

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Interactions between visual attention and episodic retrieval: Dissociable contributions of parietal regions during gist-based false recognition. *Neuron, 75,* 1122–1134. http://dx.doi.org/10.1016/j.neuron.2012.08.020

Hollingworth, A., & Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition, 113,* 150–166. http://dx.doi.org/10.1016/j.cognition.2009.08.004

Isik, L., Leibo, J. Z., & Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in Computational Neuroscience, 6,* 37. http://dx.doi.org/10.3389/fncom.2012.00037

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219. http://dx.doi.org/10.1016/0010-0285(92)90007-O

Kawachi, Y., & Gyoba, J. (2006). A new response-time measure of object persistence in the tunnel effect. *Acta Psychologica, 123,* 73–90. http://dx.doi.org/10.1016/j.actpsy.2006.04.003

Kim, J., & Yassa, M. A. (2013). Assessing recollection and familiarity of similar lures in a behavioral pattern separation task. *Hippocampus, 23,* 287–294. http://dx.doi.org/10.1002/hipo.22087

Kirchoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *The Journal of Neuroscience, 20,* 6173–6180.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science, 321,* 1502–1507.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621. http://dx.doi.org/10.1146/annurev.ne.19.030196.003045

Loiotile, R. E., & Courtney, S. M. (2015). A signal detection theory analysis of behavioral pattern separation paradigms. *Learning & Memory, 22,* 364–369. http://dx.doi.org/10.1101/lm.038141.115

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review, 14,* 1199–1204. http://dx.doi.org/10.3758/BF03193113

Moore, C. M., Stephens, T., & Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin & Review, 17,* 731–736. http://dx.doi.org/10.3758/PBR.17.5.731

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 17140–17145. http://dx.doi.org/10.1073/pnas.0914056107

Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception & Psychophysics, 67,* 324–334. http://dx.doi.org/10.3758/BF03206495

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. http://dx.doi.org/10.1163/156856897X00366

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology, 4,* e27. http://dx.doi.org/10.1371/journal.pcbi.0040027

Poth, C. H., Herwig, A., & Schneider, W. X. (2015). Breaking object correspondence across saccadic eye movements deteriorates object recognition. *Frontiers in Systems Neuroscience, 9,* 176. http://dx.doi.org/10.3389/fnsys.2015.00176

Poth, C. H., & Schneider, W. X. (2016). Breaking object correspondence across saccades impairs object recognition: The role of color and luminance. *Journal of Vision, 16,* 1. http://dx.doi.org/10.1167/16.11.1

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522. http://dx.doi.org/10.1037/0278-7393.2.5.509

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535. http://dx.doi.org/10.1037/0033-295X.99.3.518

Rentz, D. M., Parra Rodriguez, M. A., Amariglio, R., Stern, Y., Sperling, R., & Ferris, S. (2013). Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: A selective review. *Alzheimer's Research & Therapy, 5,* 58. http://dx.doi.org/10.1186/alzrt222

Richard, A. M., Luck, S. J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition, 109,* 66–88. http://dx.doi.org/10.1016/j.cognition.2008.07.004

Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: Two fundamental challenges for visual processing. *Current Opinion in Neurobiology, 20,* 382–388. http://dx.doi.org/10.1016/j.conb.2010.04.013

Scholl, B. J., & Flombaum, J. I. (2010). Object persistence. In B. Goldstein (Ed.), *Encyclopedia of perception* (Vol. 2, pp. 653–657). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412972000.n210

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38,* 259–290. http://dx.doi.org/10.1006/cogp.1998.0698

Schurgin, M. W., Reagh, Z. M., Yassa, M. A., & Flombaum, J. I. (2013). Spatiotemporal continuity alters long-term memory representation of objects. *Visual Cognition, 21,* 715–718. http://dx.doi.org/10.1080/13506285.2013.844969

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior, 6,* 156–163. http://dx.doi.org/10.1016/S0022-5371(67)80067-7

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13,* 113–142. http://dx.doi.org/10.1111/j.2044-835X.1995.tb00669.x

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10,* 89–96. http://dx.doi.org/10.1111/j.1467-7687.2007.00569.x

Stahl, A. E., & Feigenson, L. (2015). Cognitive development. Observing the unexpected enhances infants' learning and exploration. *Science, 348,* 91–94. http://dx.doi.org/10.1126/science.aaa3799

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19,* 73–74. http://dx.doi.org/10.3758/BF03337426

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia, 51,* 2442–2449. http://dx.doi.org/10.1016/j.neuropsychologia.2012.12.014

Strickland, B., & Scholl, B. J. (2015). Visual perception involves event-type representations: The case of containment versus occlusion. *Journal of Experimental Psychology: General, 144,* 570–580. http://dx.doi.org/10.1037/a0037750

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522. http://dx.doi.org/10.1038/381520a0

Van Marle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science, 14,* 498–504.

Wallis, G. (2002). The role of object motion in forging long-term representations of objects. *Visual Cognition, 9,* 233–247. http://dx.doi.org/10.1080/13506280143000412

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 4800–4804. http://dx.doi.org/10.1073/pnas.071028598

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford, UK: Oxford University Press.

Wilcox, T., & Baillargeon, R. (1998a). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology, 37,* 97–155. http://dx.doi.org/10.1006/cogp.1998.0690

Wilcox, T., & Baillargeon, R. (1998b). Object individuation in young infants: Further evidence with an event-monitoring paradigm. *Developmental Science, 1,* 127–142. http://dx.doi.org/10.1111/1467-7687.00019

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Düzel, E. (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45,* 459–467. http://dx.doi.org/10.1016/j.neuron.2005.01.010

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152–176. http://dx.doi.org/10.1037/0033-295X.114.1.152

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30,* 111–153. http://dx.doi.org/10.1006/cogp.1996.0005

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34,* 515–525. http://dx.doi.org/10.1016/j.tins.2011.06.006

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 of the United States of America, 105,* 8840–8845. http://dx.doi.org/10.1073/pnas.0802525105

Received May 31, 2016
Revision received December 8, 2016
Accepted December 9, 2016 ■