American Psychological Association

Journal of Experimental Psychology: General

0096-3445/18/$12.00

2018, Vol. 147, No. 9, 1295–1308
http://dx.doi.org/10.1037/xge0000409

Abstract Shape Representation in Human Visual Perception

Nicholas Baker and Philip J. Kellman
University of California, Los Angeles

The ability to form shape representations from visual input is crucial to perception, thought, and action. Perceived shape is abstract, as evidenced when we can see a contour specified only by discrete dots, when a cloud appears to resemble a fish, or when we match shapes across transformations of scale and orientation. Surprisingly little is known about the formation of abstract shape representations in biological vision. We report experiments that demonstrate the existence of abstract shape representations in visual perception and identify the time course of their formation. In Experiment 1, we varied stimulus exposure time in a task that required abstract shape and found that it emerges about 100 ms after stimulus onset. The results also showed that abstract shape representations are invariant across certain transformations and that they can be recovered from spatially separated dots. Experiment 2 found that encoding of basic visual features, such as dot locations, occurs during the first 30 ms after stimulus onset, indicating that shape representations require processing time beyond that needed to extract spatial features. Experiment 3 used a convergent method to confirm the timing and importance of abstract shape representations. Given sufficient time, shape representations form automatically and obligatorily, affecting performance even in a task in which neither instructions nor accurate responding involved shape. These results provide evidence for the existence, emergence, and functional importance of abstract shape representations in visual perception. We contrast these results with "deep learning" systems and with proposals that deny the importance of abstract representations in human perception and cognition.

Keywords: vision, perception, object recognition

An object's shape is a property crucial to its identity and function, and perception of shape is accordingly one of our most important capabilities. In human perception, the visual sense provides the most efficient and detailed information about shape. As a result, perception and representation of object shape through vision are basic to thought, action, and learning.

Shape is complex, however. Because shape can be described at different levels and in different ways, understanding shape perception is an enterprise that involves properties of objects but also properties of mind. The ways in which we perceive and represent shape are a subset of all possible information encoding schemes, and they are not well understood. Shape representations must capture ecologically important similarities among objects and allow classification of natural kinds despite variations (Kellman, Garrigan, & Erlikhman, 2013). Such representations must also be obtainable despite variations in viewing conditions and contexts. Although, to be useful, our shape representations must depend on relevant properties of physical objects, they are not simple or literal reflections of object properties.

The Gestalt psychologists (e.g., Koffka, 1935) were among the first to ponder deeply the nature of shape in psychological and physiological processes. Shape, as represented in the brain, is different from the collection of stimulating elements (Koffka, 1935); it depends both on the stimulus input but also on organizing activity in neural processes (Köhler, 1929). Shape is abstract: What similar shapes have in common is not their constituent elements, but the spatial relations of the parts.

Even today, how we perceive and represent abstract shape is not well understood. Work in cognitive science and neuroscience offers a variety of foundations and clues, but relatively little work has addressed the processing and representation of abstract shape. In the earliest stages of cortical visual processing, neural units register retinal regions of oriented contrast (Hubel & Wiesel, 1968; Zhang & von der Heydt, 2010). The collection of these neural responses, by themselves, does not comprise shape, nor could these initial encodings support matches of shape across transformations, such as size or orientation, or allow observers to recognize the same shape made from differing local elements.

Neural evidence suggests that abstract shape processing likely occurs in later visual areas. Single-cell recording of V4 in rhesus monkeys has found cell populations that are sensitive to shape features, such as curvature, convexity, and sharpness (Pasupathy & Connor, 2001). V4 neurons have also been found to have positional invariance, showing similar patterns of activation regardless of a stimulus's location in visual field (Gallant, Connor, Rakshit, Lewis & Van Essen, 1996). The inferior temporal cortex is also implicated in abstract shape processing. Cell populations in anterior IT remain sensitive to certain shapes, even when the size and positions of those shapes are modified (Ito, Tamura, Fujita, & Tanaka, 1995).

In cognitive science and computer vision, researchers have sought formal descriptions suitable for shape representation. Some

This article was published Online First April 9, 2018.

Nicholas Baker and Philip J. Kellman, Department of Psychology, University of California, Los Angeles.

Correspondence concerning this article should be addressed to Nicholas Baker, Department of Psychology, University of California, Franz Hall, 502 Portola Plaza, Los Angeles, CA 90095. E-mail: nbaker9@ucla.edu

have proposed that 2-D shapes are represented as abstract skeletons whose axial branches are formed based on symmetries within the shape (Blum & Nagel, 1978; Feldman & Singh, 2006; Feldman et al., 2013; Sebastian & Kimia, 2005). Another line of research has proposed that contour shape may be represented by sets of constant-curvature segments (Garrigan & Kellman, 2011).

In contrast to these neural and formal efforts, rather little work has addressed abstract shape in human perceptual processing. Dating back to the Gestalt psychologists (e.g., Koffka, 1935; Wertheimer, 1923), there are intuitive demonstrations for the reality and importance of shape as something more than the encoding of local stimulus elements. Hochberg (1968) advanced the idea that abstract "schematic maps" are synthesized from successive fixations in scene perception. However, little is known about how and when abstract representations form. In the present work, we focus on abstract shape representations, seeking clear psychophysical evidence for their existence and the time course of their formation.

The issue of encoding shape abstractly is especially timely, we believe, because of recent developments and trends in a number of fields, including cognitive science, neuroscience, and artificial intelligence. Perhaps because of the difficulty in understanding issues of how structure, such as object shape, can be extracted by perceptual systems and represented, approaches in artificial intelligence in recent years have often omitted explicit concepts of structure or shape in favor of using elaborate statistical approaches to perform object classification tasks (cf. Chomsky, 2012). Tremendous progress has been made in getting artificial systems to correctly identify objects present in photographs (e.g., He, Zhang, Ren, & Sun, 2016; Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014). Recent advances in "deep learning" systems have received a great deal of attention, both scientifically and in the popular media. It is unclear, however, to what degree, if any, such systems performing object classification make use of shape information. Although one might expect that recognizing whether a scene has a rabbit in it would involve segmentation processes that distinguish the rabbit from the background and construct a shape description that is matched to shape information about rabbits, that would not be a correct characterization of the most successful approaches. Deep learning systems, or more specifically, deep convolutional neural networks (DCNNs), process image details in a very large number of layers and at different scales, intermixed with smoothing or noise filtering operations. Encoding object shape is not a deliberate goal of such approaches, and the combination of filtering operations and training used in these systems may not lead to encoding of shape at all. As an example, Zhu, Xie, and Yuille (2016) found that DCNNs trained on natural images performed worse when tested on images with reduced backgrounds, although the target object remained intact. Conversely, these systems could classify images well above chance performance even when the target object had been fully removed from the scene.

DCNNs do not explicitly represent object shape. Whether they implicitly capture some shape properties or whether they are formally incapable of doing so in their present form is an issue of current investigation. We recently tested VGG, a popular and high performing convolutional network trained for object recognition (Simonyan & Zisserman, 2014), on glass ornaments whose abstracted shape matched an animal from one of the network's trained object categories. Figure 1 shows an example. The object shown is readily classified by human observers, based on its shape. VGG's top 5 classification responses for this object, from most to least probable, were: "hour glass," "ladle," "can opener," "loupe," and "wash basin."

Other pictures made of glass, as well as silhouettes and outlines of objects meet a similar fate (Baker, Lu, Erlikhman, & Kellman, 2018). What is the issue here? Why is the glass bunny obvious to human observers but is classified as a can opener by an artificial system (that achieves 92% performance on image sets used to evaluate systems in computer vision)? Convolutional neural networks not only use context but seem to make especially strong use of texture information, which would be expected based on the nature of the convolution operations that extract image information from local patches. What appears to be missing is a representation of abstract shape. For human observers, a bunny made of glass is surely unlikely to hop into your garden and is obviously lacking normal bunny surface texture, but the shape, even when cast in glass, is readily extracted and used for classification. One might go so far as to say that abstract shape information dominates human classification. After all, if texture were deemed most important in the classification responses, this object would not be labeled as a rabbit. The immediate and spontaneous recognition of the object's identity based on shape suggests fundamental differences between object recognition in humans and current artificial systems. This brief discussion is not meant to be the final word in comparing human and artificial vision systems, as there are issues relating to tasks, training, and so forth of the latter that are not the focus here. (For a more detailed treatment, see Baker et al., 2018.) Rather, we highlight the understanding of abstract shape perception and representation in humans as important both for understanding how

Figure 1. An example of an object whose shape is readily classified by human observers but not by deep convolutional neural networks (DCNNs); see text. (From Baker, Lu, Erlikhman & Kellman, 2018).

biological systems encode and classify objects, as well as for comparing the capabilities and limits of human and artificial systems. Further, we believe the limitations of artificial systems regarding shape processing derive in large part from our current lack of understanding of abstract shape perception and representation. Improving our understanding of shape abstraction in biological vision may offer ideas for enhancing future artificial systems. After considering the results of several experiments, we return to these issues in the General Discussion.

Within psychology, cognitive science, and neuroscience it also seems crucial to define and clarify the role of abstract shape. Somewhat paralleling trends in artificial intelligence, some recent and influential proposals have suggested that we do not really have abstract representations in perception or cognition. Barsalou (1999, 2003) argued against the existence of abstract representations in his proposals regarding perceptual symbol systems (PSS). The PSS hypothesis is that there really are no abstract representations in the usual sense; rather what has been considered as such really consists of nonabstract "re-enactment" or "simulation" of sets of basic sensory features that are activated when we perceive (Barsalou, 1999, 2003). Thus, abstract concepts "are perceptual, being grounded in temporally extended simulations of external and internal events" (Barsalou, 1999, p. 603). More concretely, as Barsalou (2003) put it:

> The basic idea behind this mechanism is that association areas in the brain capture modality-specific states during perception and action, and then reinstate them later to represent knowledge. When a physical entity or event is perceived, it activates feature detectors in the relevant modality-specific areas. During visual processing of a car, for example, populations of neurons fire for edges, vertices and planar surfaces, whereas others fire for orientation, color and movement. The total pattern of activation over this hierarchically organized distributed system represents the entity in vision(e.g., Zeki, 1993; Palmer, 1999). Similar distributions of activation on other modalities represent how the entity feels and sounds, and the actions performed on it. (Barsalou, 2003, p. 1179)

These ideas have much in common with those of classical empiricist philosophers, such as Locke, who believed that complex ideas in perception and cognition were the products of associative combination of basic sensations (for discussion, see Kellman & Arterberry, 2000; Kellman & Massey, 2013). They have been criticized for failing to offer a coherent account of abstract ideas in cognition (e.g., Landau, 1999; Ohlsson, 1999) as well as for failing understand the abstract nature of perception (Kellman & Massey, 2013). In fact, Gestalt discussions from almost a century ago (Koffka, 1935; Wertheimer, 1923) provided compelling arguments against the idea that perception could be understood as collections of sensory activations. Instead, understanding the character of perception, in general, and shape, in particular, requires abstraction, as in the classic adage that "the whole is different from the sum of its parts" (Koffka, 1935).

Our immediate purpose, however, is not to explore these issues in depth but to recognize that understanding of abstraction in perception is important for general views of perception and cognition. Much of the trend in several fields in omitting or dismissing abstract representations stems from our relatively poor understanding of how these actually work (cf., Barsalou, 2003, on difficulties with the notion of abstraction in cognitive science). One aim of the present work is to show psychophysically clear evidence of abstract shape representations, how they are processed, and their functional importance. Such efforts will hopefully lead to progress in understanding perception and abstraction in biological as well as artificial systems.

Probing abstract representations of shape requires special stimuli. A commonplace yet remarkable fact is that humans readily perceive shape from arrangements of dots. In Figure 2, the spaced, disconnected dots appear to specify a closed contour with a determinate 2D shape. No closed shape is given in the display itself, and many possible contours could connect the dots. We leveraged displays of this sort to probe the brain's formation of abstract representations of shape, as the stimuli themselves do not contain connected contours or shape features. We used patterns of black and white dots, positioned along the contours of randomly generated virtual objects (see Figure 2). Displays of this kind also allowed us to manipulate the constituent elements of the display without changing the global shape percept. Displays based on groupings of dots have been used by other investigators to explore a variety of grouping and detection phenomena (Lezama, 2015; Pizlo, Salach-Golyska, & Rosenfeld, 1997; Sha'ashua & Ullman, 1988; Smits & Vos, 1987; Uttal, 1973).

The present work focused on the abstract 2D shape representations extracted from such displays. To isolate abstract shape, we used tasks requiring comparison of shapes extracted from dot patterns across transformations of position, scale, and orientation. If abstract shape representations exist, and if an observer extracts a certain abstract shape from a dot array, s/he should be able to judge accurately whether a different dot array has the same shape, even if the second array contains a scaled or rotated version of the shape. Experiment 1 tested whether such judgments are possible and measured the processing time needed to form a shape representation that supports comparison across changes to local features and rigid 2D transformations. Experiment 2 tested whether abstract representation of shapes requires processing time over and above the time needed to register local features. In Experiment 3, we used a different, convergent method to show the existence and function of abstract shape representations. Subjects were tasked with comparing the spatial positions of dot patterns shown in sequence, with no reference to shape. When dots changed position, they did so in a way that either altered the global shape outline or left it the same. Whereas accurate registration of local features would facilitate

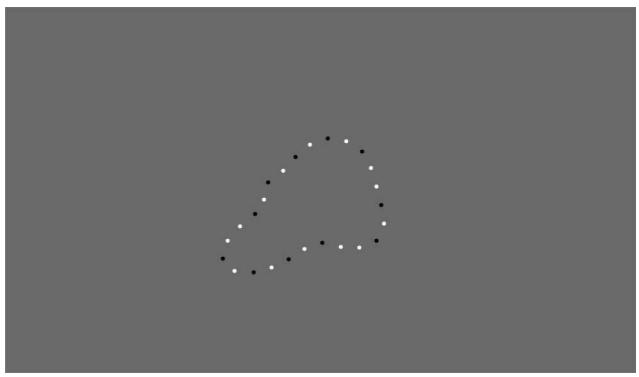

Figure 2. Example of the displays used in Experiment 1.

performance on this task, abstract shape representations might make it worse, in that detection of dot position change might be overshadowed by the formation of obligatory abstract shape representations.

Experiment 1

In Experiment 1, we tested subjects' ability to determine if the shape outlines formed by two dot patterns are the same or different across a range of encoding times. We presented subjects one novel shape for a varied duration, followed by a mask and a second shape. The second shape could differ from the first both in global outline and in position, size, or orientation on the screen. Subjects were instructed to report shapes as different only if the second shape had a different global outline than the first.

Method

Participants. Twenty-five (21 female, four male, M_{age} = 20.2) undergraduates from the University of California, Los Angeles participated in Experiment 1 for course credit. All participating subjects had normal or corrected to normal vision.

Displays and apparatus. Novel amoeba-like shapes were generated for each trial. The displays contained no continuous contours that might give shape information. Displays were comprised of 25 black and white dots evenly sampled along the contour and were displayed on a gray background screen (see Appendix A for more information).

Subjects were seated 71 cm from the 20-in. View Sonic Graphic Series G225f monitor. The monitor was set to 1024 × 768 resolution, with a refresh rate of 100 Hz.

The first display was presented at the center of the screen and subtended up to 13.8 degrees of visual angle from the most extreme left dot to the most extreme right dot (mean horizontal length was 8.00 degrees). The second shape subtended up to 18.43 degrees of visual angle (mean horizontal length of 8.03 degrees). (See Appendix B for more information).

Except when noted otherwise, all aspects of the displays and apparatus in subsequent experiments were the same as in Experiment 1.

Design. On each trial, two dot patterns were shown sequentially, separated by a pattern mask. After the second pattern was shown, subjects were asked if the second pattern had the same shape as the first pattern. Nine presentation durations—30, 50, 70, 90, 110, 130, 150, 250, and 400 ms—for the first display were presented in separate blocks of 40 trials each in a within-subjects design. Subjects completed five practice trials with feedback in which the first stimulus was presented for 500 ms and then began the official experiment, where they received no feedback.

Procedure. Each trial began with a fixation cross for 300 ms in the location of the first pattern, followed by a presentation of the first pattern for a given duration (30–400 ms), which was in turn followed by a mask of random dots for 50 ms. Following the mask, a second shape was shown. The second shape could be the same as or different from the first shape. Different shapes were generated by taking the first shape and deforming its global outline (see Appendix B). The second shape also underwent some transformation, regardless of whether or not its shape outline was altered. There were four possible conditions for the transformation of the second shape: rotation (5 to 20 degrees in either direction), scaling (between .5 and 1.5 times original shape size), translation (up to 150 pixels in any direction), and no transformation. Dot patterns were transformed in these ways to ensure that success on the task required comparisons between abstract shapes.

The second shape was always shown for 1,000 ms, and was followed by another mask for 300 ms. Subjects performed a forced choice same/different task. They were instructed to report "Same" if the two dot patterns had the same shape outline and to report "Different" if the second pattern had a different shape outline, irrespective of the rigid body 2D transformation. See Figure 3 below for a sample trial of Experiment 1.

Dependent measures and data analysis. We measured subjects' accuracy on the same/different task across the nine presentation times for the first display. Data were analyzed by taking each subject's average performance for each of the nine presentation times, and then computing a group average and confidence intervals. Performance was statistically compared across the several exposure durations and to chance performance. To eliminate possible effects of bias from subjects tending to say "same" or "different," we also used signal detection methods to measure sensitivity (d') as a function of encoding time. Finally, we used logistic and piecewise regression analyses to identify the encoding time beyond which a stable, abstract shape representation was available (see below).

Results

Figure 4 shows the mean accuracy data for the 9 exposure durations for the first display. Performance was better than chance in all conditions except at the 30 ms exposure duration (all ts > 2.99; all ps < .01); at 30 ms, the mean accuracy of .508 (95% confidence interval [.482,.533]) did not differ from chance, $t(24) = .64$, $p > .250$.

Performance improved with encoding time, up to 110 ms, after which it plateaued. To identify the point at which more processing time ceased to produce improvements in the comparison task, we fit the results to a psychometric function using the Palamedes Toolbox (Prins & Kingdom, 2009). The maximum likelihood estima-

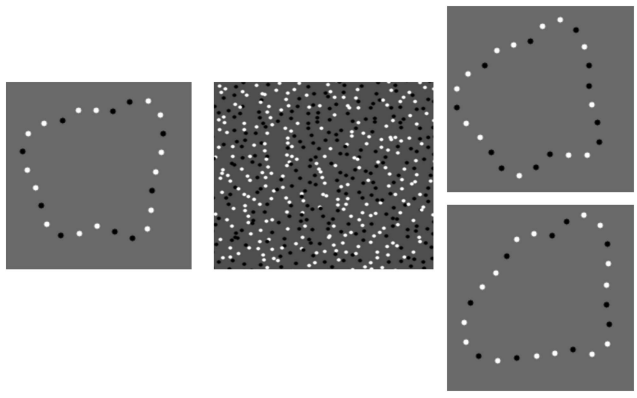

Figure 3. Sample trial for Experiment 1. The first display is on the left, followed by a pattern mask. The second display could either have the same shape as the first with some transformation (top right) or the shape could be deformed in some way (bottom right).

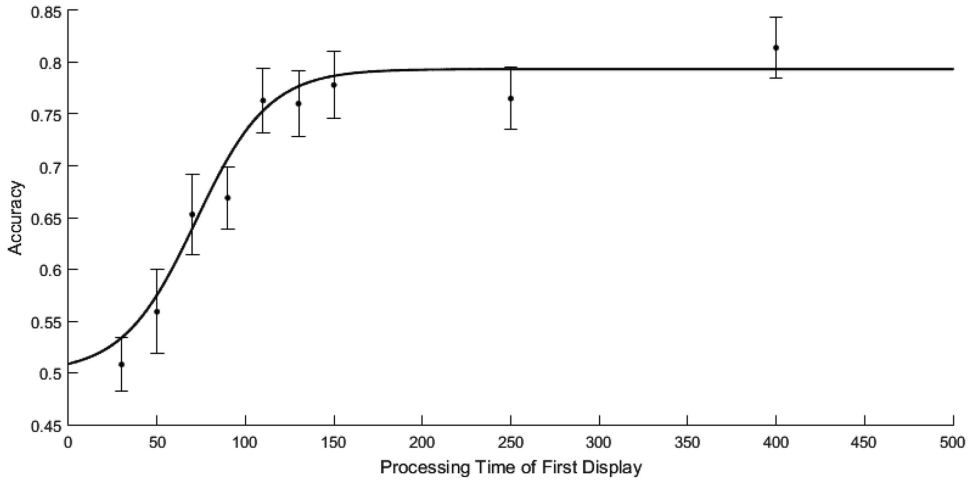

Figure 4. Accuracy as a function of exposure duration in Experiment 1. Error bars show 95% confidence intervals.

tion for the data is given by the function $.5 + \frac{.293}{1+e^{-48.476(x-.072)}}$, where x is the amount of processing time in seconds. By taking the second derivative of this function, we identified the point at which performance flattened. In these data, this transition point was at 99.1 ms. As another way of identifying this transition point, we used a continuous piecewise regression, with a change of slope at one of the experimental viewing durations between 50 and 250 ms. We compared the R^2 value for each of these seven regressions in order to determine which possible transition point explained the most variance in our data. R^2 was highest (.617) for a piecewise regression whose transition point was at 110 ms, $F(2, 222) = 178.60$, $p < .001$. There was a reliable difference in slope for observations between 30 and 110 ms and for observations between 110 and 400 ms, $t(2) = -12.46$, $p < .001$. The piecewise regression gave a predicted gain in accuracy of 3% per 10 ms before 110 ms, and less than .2% per 10 ms beyond the inflection point. Piecewise regressions with transitions at other points were significant, but accounted for less of the overall variance.

Results were also analyzed using a signal detection theory measure of sensitivity, with a correct detection of a change being considered a hit and an incorrect change response being a false alarm. The results are shown in Figure 5. The pattern of results was almost identical to the data with accuracy in Figure 4, with performance leveling off beyond 110 ms of encoding time.

Individual transformations. The design of Experiment 1 aimed at determining the time course for the construction of abstract shape representations and tested for them by requiring shape comparison across transformations. Our hypothesis was that successful performance for all transformation types, with the possible exception of translation, would require an abstract shape representation. Alternatively, it could be the case that the visual system actually uses differing underlying representations to com-

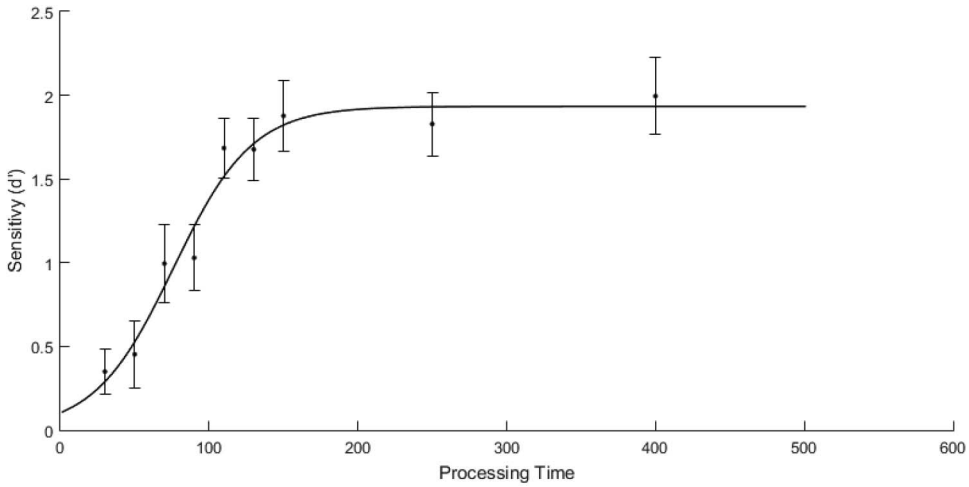

Figure 5. Sensitivity (d') as a function of exposure duration in Experiment 1. Error bars show 95% confidence intervals.

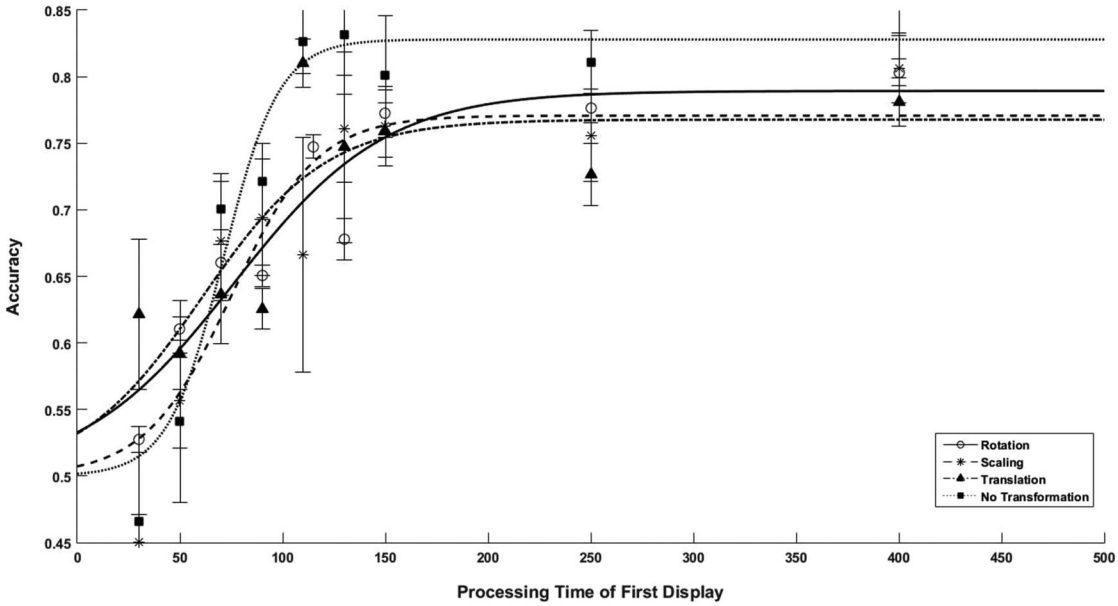

Figure 6. Accuracy as a function of exposure duration for separate transformations in Experiment 1**.** Error bars show 95% confidence intervals.

pare displays that differ by translation, scaling, and rotation, in which case performance as a function of exposure duration might differ across transformation. To assess these possibilities, we examined each of the rigid 2D transformations separately. Figure 6 shows these results. Using the continuous piecewise regression test described above, we looked for a transition point in each of the transformations, looking for the regression with the highest R^2 value. Results are shown in Table 1. R^2 was highest when the transition point was fixed at 110 ms for two of the three transformations. Notably, even in trials where no 2D transformation was introduced, the data are best explained with a transition point at 110 ms. When the shape was scaled, a change of slope at 70 ms of encoding time proved to explain the variance better than at 110 ms (see Table 1). The earlier transition point identified for scaled shapes is likely a statistical anomaly, driven by a particularly low mean accuracy at 110 ms. This is also supported by the second derivative of the logistic function test, which identified a transition point at 102 ms. Apart from this outlier, the trajectory for scaled transformation mirrors the other three conditions.

Table 1
Transition Point Estimated Using Logistic and Piecewise Linear Regression by Condition in Experiment 1

Condition	Transition point (ms)	
	Logistic regression	Piecewise linear regression
All transformations	99.1	110
Rotation	124.8	110
Dilation	102.0	70
Translation	100.0	110
No transformation	88.8	110

Discussion

From 30 to 110 ms of viewing time for the first shape, subjects go from being unable to compare virtual shapes at better than chance accuracy to achieving a consistent high level of discrimination between displays of the same or different shapes. Presentation of a mask between the first and second display in this study should prevent subjects from using apparent motion (Braddick, 1973) or visual icons (Sligte, Scholte, & Lamme, 2008; Smithson & Mollon, 2006) in the comparison task. Because of the transformations used, encoding and comparison of local elements (dots) from each display would not support accurate performance. These arrangements aimed to require subjects to compare displays based on an abstract representation of a global contour connecting the dots in each display. That results were similar across differing transformation types, and even for the no-transformation condition, suggests that a common abstract shape representation was used in the task. This representation does not appear to be available at the shortest presentation time tested (30 ms) but appears to be fully available by about 110 ms. This contour must be constructed in an object-centric, not retinotopic, format to make shape comparison possible across 2D transformations.

An open question these data raise is what is happening between 30 and 110 ms of processing time. One possibility is that viewers have access to partially formed representations of shape during the time between when encoding begins and when it is completed. Another possibility is that shape representation is discrete but probabilistic. Under this hypothesis, comparison between shapes can only be accurately carried out when a complete abstract representation has formed, but there is a distribution over the time this formation requires, with rather low probability at 50 ms of processing time, and very high probability with more than 110 ms.

The results suggest that a set of dots arranged along the contour of a virtual shape produces an abstract shape representation. Our task was designed to require comparison of abstract shape, and participants' similar performance across the transformation types we tested is consistent with the use of such a representation. The results of Exp. 1 suggest that abstract shape representations are not immediately available from a display but require on the order of 110 ms to be fully formed.

Experiment 2

Experiment 1 showed that subjects could not produce their best performance in comparing two dot patterns' virtual contours with less than 110 ms of processing time. The most natural explanation is that the task required abstract representations not explicit in the physical stimulus, and such representations take measurable time to be constructed, beyond the time needed to register the physically given elements in the display. Another possibility, however, is that abstract shape becomes available as early as basic stimulus elements (dots) are encoded. The time course we measured may simply reflect the time required for sensory encoding of the stimulus elements. If this is the case, the results of Experiment 1 would have little to do with time constraints on abstract representations of shape per se. We test this possibility in Experiment 2 by testing whether visual features are adequately registered even at the shortest encoding time used in Experiment 1 (30 ms).

A variety of work in vision suggests that encoding of basic features happens substantially faster than 110 ms (Ringach, Hawken, & Shapley, 1997; Subramaniam, Biederman, & Madigan, 2000). Making a rigorous claim psychophysically about encoding time for basic features is difficult, however. If no pattern mask is used, processing of a display shown briefly will continue after the stimulus is removed (Schultz & Eriksen, 1977; Sperling, 1960). In that case, it is hard to make the claim that basic features were registered within the display interval. Conversely, use of a pattern mask halts processing but also tends to obliterate any records of local features. Therefore, a task aimed at explicitly assessing encoding of dots in specific locations, and using a pattern mask, would reveal little in the way of such records, unless there is time to recode elements into a more durable store (Coltheart, 1980; Sperling, 1963, 1967). To avoid these two difficulties, we used an indirect task, using detection of transformations between briefly presented displays, where the second display might function as a mask, but motion mechanisms might still reveal specific encoding of the original elements. If subjects could succeed in classifying transformations of dot patterns that depended on the spatial positions of the initial elements, it would provide evidence that abstract shape representations involve time demands beyond those needed for basic encoding of stimulus elements.

Method

Participants. Twenty-six (23 female, three male, M_{age} = 20.0) subjects participated in Experiment 2. Subjects were undergraduates from the University of California, Los Angeles with normal or corrected to normal vision who earned extra credit for participating in the study.

Design. The experiment included 200 trials, including 100 rotational transformations (half clockwise, half counterclockwise), and 100 scaling transformations (half larger, half smaller). Subjects completed five practice trials with feedback to ensure they understood the task before the main experiment began.

Procedure. Novel dot patterns generated the same way as in Experiment 1 were displayed for 30 ms. Following a 10-ms interstimulus interval with a blank screen, a second display was shown. The second display had the same set of dots as the first, but the set was either rotated clockwise or counterclockwise 10 to 25 degrees, or it was scaled by a factor between 1.2 and 1.45 when enlarged and between 1/1.45 and 1/1.2 when made smaller. Subjects were first asked whether the dot display was rotated or scaled, and based on their response, they were asked the direction of the transformation (clockwise or counterclockwise if subjects answered "rotated," and larger or smaller if subjects answered "scaled"). Trials were scored correct only if subjects correctly answered both questions. Note that although the first and second displays had the same abstract shape description in all cases, this task can be done via apparent motion mechanisms operating on the individual dot elements, without computation of global shape tokens (Ullman, 1979).

Results

One subject's data was removed because her performance was more than three standard deviations from mean performance. Analysis was carried out with and without her data with no meaningful differences. Figure 7 shows the primary data from this experiment, along with the 30-ms exposure duration condition from Experiment 1. Mean accuracy for Experiment 2 was .93, 95% confidence interval (CI) [.913, .944]. Performance in Experiment 2 was reliably higher than chance, $t(6) = 149.30$, $p < .001$, and significantly different from performance in Experiment 1, $t(30) = 18.15$, $p < .001$.

Discussion

This experiment sought evidence that basic feature registration (dots and their positions) could be accomplished even at the

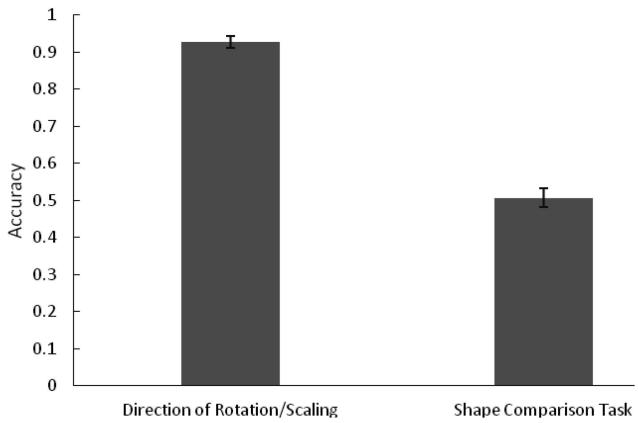

Figure 7. Accuracy on the dot transformation task in Experiment 2. The bar on the left shows accuracy for clockwise and counterclockwise rotation of the set of dot elements, and for expanding or contracting scaling. Accuracy on the abstract shape task of Experiment 1 for the same (30 ms) exposure duration is shown on the right for comparison. Error bars show 95% confidence intervals.

shortest interval tested in Experiment 1, substantially less than the 110 ms required to perform a task based on abstract shape. Subjects in Experiment 2 showed nearly perfect accuracy when the initial display was presented for 30 ms. In contrast, subjects in Experiment 1 performed at chance at that exposure duration. Accurate responding in Experiment 2 required that the spatial locations were extracted from dots in the first display. If this were not the case, subjects would have no reference with which to compare dot locations in the second display, and performance would suffer. We believe the "comparison" here comes from incorporation of the two displays into basic motion computations (Dawson, 1991; Ullman, 1979), but the requirement for stimulus registration is implicit in that mechanism. Taken together, the results of Experiments 1 and 2 suggest that early visual feature registration occurs within 30 ms, but construction of abstract shape representations requires additional processing time.

Experiment 3

In Experiments 1 and 2, we found evidence for the existence of abstract shape representations by comparing the processing time needed to perform a task that required an abstract shape description (Experiment 1) with the processing time needed for the registration of physical features (Experiment 2). In Experiment 3, we used a convergent method to reveal the existence and functional effects of abstract shape representations.

Subjects were shown two dot patterns in sequence. Whereas in Experiment 1, we directed subjects' attention to shape—tasking them to determine if the two dot patterns had the same shape—in Experiment 3 we asked them to decide if any of the dots in the second display occupied a different spatial position than the dots shown in the first display. On a third of the trials, the dots did not change position at all. On another third of the trials, dots were displaced in a random direction. On the last third, dots were displaced along the existing shape contour. Subjects' assigned task was to attend to the physical positions of dots on the screen, and no mention of shape was made. Conceptually, this task requires no processing of abstract shape; ideally, it would be performed by registering exact positions of dots in the first display and detecting differences from these positions in the second display.

Although this experiment was carried out to investigate issues of abstract shape representation, it is also relevant to theories in cognitive psychology that suggest that abstract thought and representations are derived by the brain revisiting literal encodings of sensory elements (in particular, the PSS hypothesis of Barsalou (1999)). In contrast, we hypothesized that this paradigm might reveal that after brief initial processing, encoding of local elements is poor, especially when more abstract representations have been derived from them.

Method

Participants. Twenty-five (17 women, eight men, M_{age} = 20.46) undergraduates from the University of California, Los Angeles participated in Experiment 3 for course credit. All participating subjects had normal or corrected to normal vision.

Displays and apparatus. Shape contours were generated using the same algorithm as in Experiment 1 and 2, but the 25 dots for the first shape were sampled somewhat differently. Initially evenly spaced dots were randomly assigned black or white color, as in previous experiments, but in Experiment 3, the positions of the dots were jittered along the shape contour so that the distance between dots along the contour was not constant.

When the second array differed from the first, it was generated in one of two ways. One way involved evenly sampling another 25 dots from the same shape contour, but with a different starting point such that the sampled dots were at the midpoints between sampled dots from the first array. Then, the same jittering procedure was performed on the dots as in the first shape.

The other method for generating the second shape was by displacing each dot from the first shape in a random direction, with no requirement to remain along the shape contour (see Figure 8). The average distance of dot displacement was equated for both of these two methods.

To prevent subjects from adopting the strategy of saying "different" any time the second display's dots were not arranged along a virtual contour, we also included trials in which the dots in the first display did not fall along a virtual shape contour, and the second display's dot positions could either match or differ from the first's. So, if subjects encoded information from only one of the two frames, they could not reliably give the correct response based on the characteristics of the stimulus they saw. Trials in which the dots in the first display were not arranged along a shape contour were not included in the analysis.

Design. The experiment used a 2 (presentation duration) × 3 (display transformation type) factorial design, with 40 trials per

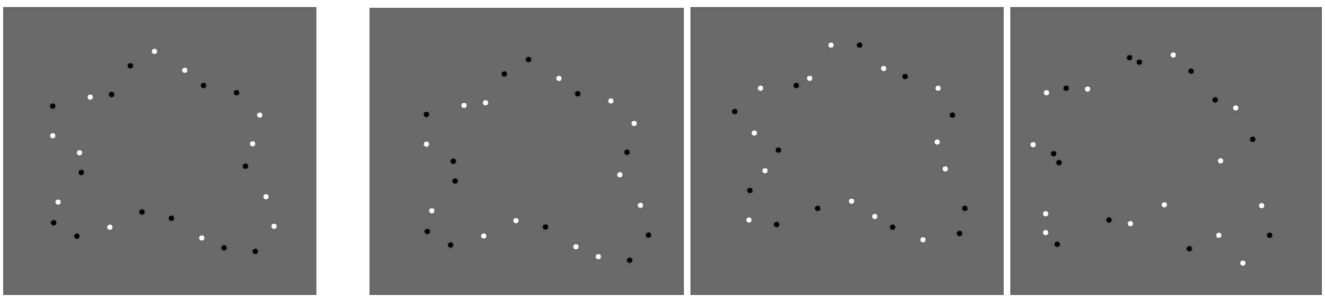

Figure 8. Sample trial from Experiment 3. On the left is the first display and on the right are three possibilities for the second display: (a) Dot positions are identical and only dot colors change, (b) dot positions have moved along the virtual shape contour, or (c) dot positions are moved in a random direction.

Table 2
Conditions for Experiment 3

Display 1	Display 2	Exposure duration	Correct response
25 dots, along a shape contour	Same 25 dots along the shape contour	30 ms	Same
25 dots, along a shape contour	Different 25 dots along the same shape contour	30 ms	Different
25 dots, along a shape contour	Different 25 dots, not along the shape contour	30 ms	Different
25 dots, along a shape contour	Same 25 dots along the shape contour	150 ms	Same
25 dots, along a shape contour	Different 25 dots along the same shape contour	150 ms	Different
25 dots, along a shape contour	Different 25 dots, not along the shape contour	150 ms	Different

condition. Trial conditions are shown in Table 2. Presentation duration for the first display was 30 ms on half of the trials and 150 ms on the other half. The 30-ms duration was chosen to be briefer than needed to form an abstract shape representation, according to the results of Experiment 1, and the 150 ms duration was chosen to be longer than needed to form abstract representations. In a third of the trials, the second display was identical to the first display. In another third of the trials, the second display contained a pattern with dots moved along the shape contour; and in the last third, the second display contained a pattern with dots moved in random directions. Dot colors were always randomly reassigned in the second display.

Procedure. On each trial, two displays of black and white dots were shown, one after another. Following presentation of the second display, subjects were asked if the positions of any of the dots had changed from the first display to the second. The first display (shown for 30 ms or 150 ms) was always cued by the presentation of a fixation cross for 300 ms and was followed by a pattern mask of random black and white dots for 300 ms. The second display was shown 1,000 ms following the pattern mask. A second pattern mask was shown for 300 ms following the second display, after which subjects performed a two-alternative forced-choice task. They were instructed to say "different" if they judged that any dots had changed position in the second display. They were instructed to say "same" if they judged the dots in the second display to be in identical positions to those in the first display.

Results

Figure 9 shows the accuracy results from Experiment 3. Detection of change in dot positions appears to be roughly at chance for 30 ms exposures of the first display in all conditions. At 150 ms, performance was above chance when dots underwent random position changes or did not move. When dots moved along the virtual shape contour, however, performance was worse than chance responding. A 2 (presentation duration) × 3 (transformation condition) analysis of variance (ANOVA), with both factors within subjects was carried out to confirm these patterns. There was a reliable main effect of presentation duration, $F(1, 24) = 4.86$, $p = .004$, and a reliable main effect of transformation condition, $F(2, 24) = 11.48$, $p < .001$. There was a significant duration by transformation interaction, such that accuracy at the longer presentation duration was higher when the second display was identical to the first, or when the dots in the second display moved in a random direction, but was lower when the dots in the second display moved along the existing shape contour, $F(2, 48) = 43.31$, $p < .001$.

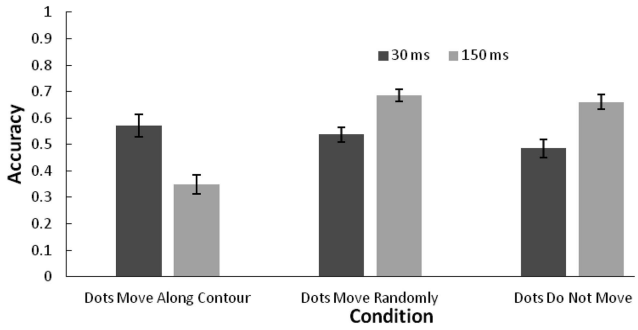

Figure 9. Accuracy data from the three transformation conditions in Experiment 3. Error bars show 95% confidence intervals.

When subjects had 30 ms to view the first display, performance was near chance regardless of the second display condition. For dots moving along the contour, subjects did reliably better than chance, with a mean accuracy of .57, 95% CI [.51, .63], whereas in the other two conditions chance performance fell within the 95% confidence intervals. These three conditions differed marginally from each other (all ts < 1.99, all ps $> .058$).

When subjects had 150 ms to view the second display, all conditions differed from chance performance. When dots in the second display were moved along the shape contour, subjects did significantly worse than chance, with a mean accuracy of .35, 95% CI [.279, .418]. When dots in the second display were moved in random directions, subjects performed significantly better than chance, having a mean accuracy of .686, 95% CI [.637, .735]. Likewise, when dots occupied the same position in the second display as in the first, subjects had a mean accuracy of .661, 95% CI [.604, .718].

The data were also analyzed using signal detection measures. We defined the signal as a change in dot positions between the two displays. On this basis, a *hit* was a trial on which an observer correctly detected a change in dot positions, whereas a *false alarm* consisted of an observer responding that dot positions had changed when in fact no dots had changed positions. The hit and false alarm rates were used to calculate sensitivity (d'), shown in Figure 10. Note that no sensitivity is given for when dots do not move because that event is defined as the absence of signal. The analysis could be framed in the reverse way (with "no change" defined as signal), in which case the d' values would remain the same. We defined the change as signal, as it seems more intuitive, allows for

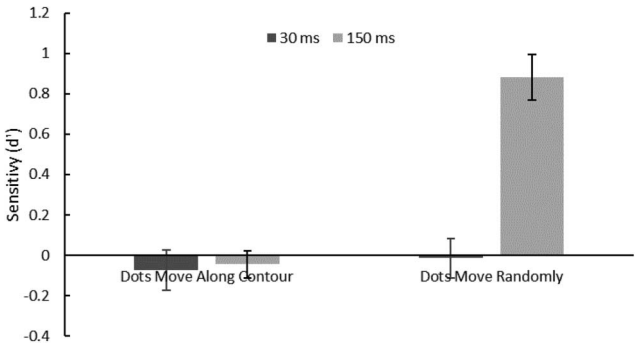

Figure 10. Sensitivity (d') data from Experiment 3. Error bars show 95% confidence intervals.

simpler condition labels, and keeps the "no-signal" case identical across the experimental conditions.

Inspection of the data indicates that observers had approximately zero sensitivity for both conditions when the duration of the first display was 30 ms, as well as in the 150-ms duration for the first display when dots moved along the contour. A 2 (condition) × 2 (presentation time) ANOVA found a significant interaction between the presentation duration of the first dot pattern and the nature of the dot movement in the second display, $F(1, 24) = 29.39$, $p < .001$. There was a reliable main effect for both presentation time, $F(1, 24) = 47.89$, $p < .001$ and type of dot movement, $F(1, 24) = 34.86$, $p < .001$. Sensitivity was not significantly different from zero in trials where the first dot pattern was shown for 30 ms and dots were moved along the contour, $t(24) = -1.225$, $p = .233$, nor when they were moved in random directions, $t(24) = -.672$, $p > .250$. When the first dot pattern was shown for 150 ms, subjects' sensitivity was not significantly different from zero when dots were moved along the contour, $t(24) = -.066$, $p > .250$, but was significantly greater than zero when dots were randomly perturbed, $t(24) = 8.604$, $p < .001$.

Discussion

In Experiment 3, subjects were instructed to attend the physical positions of dots on the screen, not to shape. No mention of shape was made when explaining the experimental task. The data suggest, however, that formation of abstract shape representations exerted an important, and obligatory, effect on subjects' performance.

When displays were exposed for 30 ms, subjects showed roughly chance accuracy and zero sensitivity to dot location changes. In terms of abstract shape representations, this was expected because such representations take longer than 30 ms to form. In terms of physical locations of elements, these would have registered within 30 ms, but may have been prevented from being encoded into a more durable representation by the pattern mask. When subjects were given sufficient processing time to form an abstract shape representation, performance on the dot position task depended closely on whether overall shape cues were congruent or incongruent with dot movements. For dots that moved randomly, altering abstract shape, subjects at 150 ms were well above chance in detecting the second display as different. For dots that retained the same position, both local position cues and overall shape remained the same, and subjects performed above chance in detecting that that first and second displays were the same. Most crucially, when dots were shifted along the contour between the first and second displays, the abstract shape representation (preserved) was incongruent with local dot positions (altered) with reference to the perceptual decision to be made. Here, subjects were reliably lower than chance in reporting whether the physical positions of the dots had changed and in signal detection analyses showed zero sensitivity for detecting dot displacement.

These results have several implications. First, the contrast between the 30-ms and 150-ms presentation conditions provides further evidence that abstract shape representations take time beyond 30 ms to form. Specifically, the difference in conditions at 150 ms seems to be a consequence of the formation of an abstract shape representation; no such difference was found at the 30-ms presentation duration. Second, use of abstract shape representations appears to be obligatory: Despite having no role in the assigned task, whether overall shape was preserved or changed between the first and second displays appeared to dominate subjects' response patterns. Third, the results provide strong evidence that abstract shape representations are derived from, but do not consist of, sets of feature activations that were present during initial sensory registration. Despite instructions to encode local dot positions, this study provides no evidence that local dot positions were encoded into any enduring representation, even at the longer presentation duration. Although abstract shape depends on registration of sensory elements, these inputs appear to be rapidly discarded. These results have implications for some proposals about the nature of abstract representations in perception and cognition. According to the PSS hypothesis proposed by Barsalou (1999, 2003), abstract ideas are not in some special abstract representational format; rather, they involve simulating activation of early sensory areas responsible for detecting features physically present in the stimulus (see Kellman and Massey, 2013, for further discussion). Such a system would have little difficulty detecting a difference in dots displaced along a virtual shape contour, as none of these early feature activations would be the same from the first display to the second. Instead, the role of early sensory activations appears to be to allow more abstract relations to be computed, with the raw material rapidly discarded rather than encoded in an enduring way. These results reflect classic insights, by the Gestalt psychologists and others, into the relational and abstract nature of perception (Gibson, 1979; Koffka, 1935; Kellman & Massey, 2013; Michotte, Thinés, & Crabbè, 1964).

General Discussion

The purpose of this work was to investigate the psychological reality of abstract shape representations. We had two hypotheses about these representations. The first hypothesis was that if abstract representations of shape exist, their formation requires processing time beyond time needed to register the local spatial features composing the shape in a visual scene. The second hypothesis was that these shape representations may not preserve information about the spatial features from which they were originally constituted.

In Experiment 1, we measured subjects' ability to compare two shapes across a variety of 2D transformations, while limiting the processing time for the first shape. We found that when subjects

were limited to 30 ms of processing time to encode information from the first display, performance was at chance level, and improved monotonically up to 110 ms, after which more processing time produces little or no improvement.

To determine if the time needed to do the task in Experiment 1 corresponded to the time needed to extract local features from a display, we asked subjects to describe a rigid transformation on a dot display in Experiment 2. Subjects were shown a display for 30 ms, followed by another display in which the set of dots had been either rotated or scaled. Subjects were extremely accurate in reporting both the kind of transformation and its directionality, tasks which required information about the positions of dots in the first display.

In Experiment 3, we tested subjects' ability to detect changes to local elements when these changes did and did not produce a change in the shape. The results indicated little or no ability to encode local element positions sufficiently to do the task. This does not mean that durable local encoding of elements is impossible; with practice or in the absence of global shape representations, it seems likely to be possible. In this experiment, however, the presence of sameness or difference in abstract shape representations dominated responses; dot changes were only detected with nonzero sensitivity when they were accompanied by form changes.

An apparent paradox in thinking about our results relates to classic ideas from Gestalt psychology and research on Configural Superiority Effects (CSEs). Gestalt psychologists pointed out phenomena in which wholes seemed to be accessed before parts, and they often claimed that parts gain their meaning from their relation to the whole (Koffka, 1935). Research by Pomerantz and colleagues on CSEs has shown many examples in which displays are more quickly distinguished when the difference arises from certain relations between elements, rather than from physical properties of elements on their own (Pomerantz & Portillo, 2011; Pomerantz, Sager, & Stoever, 1977). For example, participants' response times are longer when they are asked to determine which of four displays is different (the "odd-quadrant" paradigm) based on the spatial position of a single dot in each quadrant than when a reference dot is added to each quadrant, giving four two-dot configurations, one of which differs in proximity or orientation from the others (see Pomerantz & Portillo, 2011, p. 1340). Why is it, then, that we find physical stimulus features are extracted more quickly than abstract relations in Experiments 1 and 2?

We posit that there is an initial, transient registration of local spatial features to which human perceivers do not have conscious access. Use in perceptual tasks requires encoding into a more durable store (Sperling, 1960). This early registration is used in the formation of abstract shape representations but can also be directly observed when it interacts with motion mechanisms, as in Experiment 2. However, once an abstract representation of shape has formed, much of the physical information from this earlier representation is lost, as was observed in Experiment 3. We believe that configural superiority effects derive from properties of more durable representations used in perceptual tasks. This view of configural effects is consistent with current understandings of early cortical processing, in which information comes in through highly local receptive fields. Initial sensory registration to which perceptual processes have access (as in our motion task, or if cued in a Sperling-type task) rapidly decays if not recoded. Most or all Gestalt effects happen beyond this early registration and presuppose it. Our experiments indicate that abstract shape, which is not in the stimulus per se, takes a certain amount of time to be acquired, and that is longer than would be required for initial sensory registration. CSEs, we believe, are probably effects showing configural priority within more stable representations. What may be surprising about this interpretation is that perceptual representations are abstract encodings synthesized from initial sensory registration, even for something as simple as a single dot. (See Kellman & Massey, 2013 for similar arguments regarding perception of apparently simple properties such as edge orientation and color.) We believe even the CSE tasks that hinge on the positions of single dots in each quadrant are operating on a postsensory representation. What is a bit counterintuitive in this explanation is that, although the odd-quadrant discrimination task in the single-dot case could in theory be done based on local spatial information in the initial sensory registration, it is not done that way; rather the positional information must be gleaned from the more enduring representation. In the latter, as the Gestaltists emphasized and as CSEs show, relational properties are of higher priority, whereas the exact coordinates of a dot are not well represented. As we know from induced motion studies in the same tradition, a dot in empty space, stationary in relation to an observer, will nevertheless appear to move if a surrounding frame moves (Duncker, 1929). This idea is also consistent with the relatively poor accuracy found by Pomerantz and Portillo (2011) in single-dot conditions. The power of the reference dots (in the comparison case of two-dot configurations) is that, although by themselves they add no real information, they create configurations that are highly salient and rapidly accessed in the task.

These interpretations are also consistent with the time courses of various phenomena. Registration in V1 after the onset of a stimulus takes about 20–60 ms (Maunsell & Gibson, 1992). In the present experiments, abstract shape seems to be accessible after about 110 ms. Paradigms differ, making simple comparisons difficult, but Pomerantz and Portillo (2011) found response times in odd-quadrant CSE experiments with single dots on the order of 1,400 ms, versus about 1,100 ms for the two-dot patterns. Even if response initiation and execution components are on the order of 500 ms in this task, this leaves 600 ms or so for perception and decision.

All of these observations fit with a view that configural effects probably derive from processing carried out on more stable representations that derive from earlier, more transient, sensory ones. On this view, there is no real paradox when considering together CSE results and the present results on abstract shape perception. Although our goal in this article has been to shed light on abstract shape representations that form even in the absence of continuous contour information in the stimulus, these ideas are consistent with a more general point that we tend to underestimate the amount of abstraction that is intrinsic to ordinary visual perception (cf. Kellman & Massey, 2013).

A possible limitation of our study is that all our experiments were conducted on dots sampled from the contours of amoeba-like shapes. In these experiments, it was essential that all shapes be unfamiliar to the viewer and share many of the same local curvature features to ensure that subjects were encoding a representation of the presented shape rather than matching it to an existing template of an object's shape or encoding only a salient feature of the contour. It is easy to imagine that other kinds of shapes might be encoded abstractly with

slightly more or less processing time than was observed in Experiment 1, but we believe in all cases this processing time will be measurably longer than the time needed to extract local spatial features of the elements from which the shape is constituted.

These experiments provide relatively direct evidence that abstract shape representations exist and require meaningful processing time to form. As indicated in the discussion of Experiment 3, they are inconsistent with proposals that claim that higher level perception and cognition are based on reactivation of sensory elements or features. Instead, these results directly implicate abstract representations, which are formed from relations of sensory elements, but do not correspond to them. In Experiment 3, any system that actually recorded feature activations and could retain access to them would have led to perfect performance. Less complete registration of features would still have produced a markedly different set of results than what we observed. Our results provide evidence instead for formation of an abstract shape representation and discarding of sensory elements used as raw material in construction of such representations. Such abstraction is more likely the rule than the exception in ordinary perceptual processing, even for seemingly simple properties of objects in the world, such as orientation or color, much less shape (Garrigan & Kellman, 2008; Kellman & Massey, 2013).

The current results also have implications for understanding the relations between perception in biological systems and artificial systems that perform classification tasks. In the past 10 years, deep learning neural networks, especially convolutional neural networks (e.g., He et al., 2016; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014), have achieved previously unattained performance on image classification tasks. This remarkable success has raised questions about whether analogues exist between these trained deep networks and the visual brain (Güçlü & Gerven, 2015; Kriegeskorte, 2015; Yamins et al., 2014).

A number of considerations suggest that although the current generation of artificial systems may perform interesting computations and have utility for some tasks, they operate profoundly differently from perception in biological systems. As our results indicate, human perception of objects relies on processes that abstract shape within 110 ms or so after stimulus presentation. Such shape processing appears to provide a means of recognizing and classifying objects that allows abstraction over many other object properties. For example, as illustrated in the introduction, humans readily recognize a rabbit or elephant in a glass object, an outline, or a silhouette. Artificial systems appear to have little or no access to overall shape information, and they appear to be heavily dependent on local texture information. We draw this conclusion both from the nature of the convolution operations that underlie these systems as well as from studies of their output. In other simulations we have carried out, preservation of shape information while substituting different texture (e.g., overlaying a wolf's fur on a bear's silhouette) leads reliably to preferred classification by texture rather than outline by DCNNs (Baker et al., 2018). These observations are consistent with other recent results; for example, certain pixel changes that do not affect human recognition can result in misclassification in DCNNs (Szegedy et al., 2013; Ullman, Assif, Fetaya, & Harari, 2016). Segmentation and the encoding of an object's shape appears to play a much more critical role in human vision than in DCNNs. The latter class of systems are of great interest, but it is possible that they are in principle limited by the absence of abstract shape coding. Efforts to further understand and model abstract shape perception and representation in humans may be the key to additional major advances in artificial perceiving and classifying systems.

Abstract shape representations are real, and they are distinct from the physical elements composing a contour. The experiments reported here indicate that such representations take approximately 110 ms to form. The results also indicate the primacy of abstract representation in perception; early featural encoding supports development of more abstract and enduring representations. Shape, and other abstract relations, allow perceptual systems to capture crucial properties of objects, spatial arrangements, and events. Initial sensory registration of local features forms the basis from which abstract representations are derived, but they may typically have little enduring effect beyond that. As a consequence, abstract representations may be employed even when more literal feature records would support better task performance. This representational primacy of the abstract probably reflects the functional importance of the kinds of spatial and temporal structure that perceptual systems must capture to be most useful in thought and action.

Context of the Research

This work originates from Nicholas Baker and Philip J. Kellman's interests in structure, relations, and abstraction in visual perception. The findings relate to programmatic efforts in our research to understand the connections between the encoding of local information early in visual pathways and meaningful perceptual representations of objects, space and motion that underlie thought, action, and learning. How these connections operate, sometimes described as the linkage between subsymbolic and symbolic visual processes, encompasses some of the most fundamental unsolved problems in the psychology, cognitive science, and neuroscience of perception. In future work, we hope to continue recent efforts to understand these phenomena through experiments and modeling.

References

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). *Deep convolutional networks do not make classifications based on object shape*. Manuscript in preparation.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18,* 513–562. http://dx.doi.org/10.1080/01690960344000026

Blum, H., & Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition, 10,* 167–180. http://dx.doi.org/10.1016/0031-3203(78)90025-0

Braddick, O. (1973). The masking of apparent motion in random-dot patterns. *Vision Research, 13,* 355–369. http://dx.doi.org/10.1016/0042-6989(73)90113-2

Chomsky, N. (2012). Noam Chomsky on Where Artificial Intelligence Went Wrong: An extended conversation with the legendary linguist. *The Atlantic*. Retrieved from https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228. http://dx.doi.org/10.3758/BF03204258

Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem.

Psychological Review, 98, 569–603. http://dx.doi.org/10.1037/0033-295X.98.4.569

Duncker, K. (1929). Über induzierte Bewegung. *Psychologische Forschung, 12,* 180–259. http://dx.doi.org/10.1007/BF02409210

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 18014–18019. http://dx.doi.org/10.1073/pnas.0608811103

Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. (2013). An integrated Bayesian approach to shape representation and perceptual organization. In S. Dickinson & Z. Pizlo (Eds.), *Shape perception in human and computer vision* (pp. 55–70). London, UK: Springer. http://dx.doi.org/10.1007/978-1-4471-5195-1_4

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of neurophysiology, 76,* 2718–2739. http://dx.doi.org/10.1152/jn.1996.76.4.2718

Garrigan, P., & Kellman, P. J. (2008). Perceptual learning depends on perceptual constancy. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 2248–2253. http://dx.doi.org/10.1073/pnas.0711878105

Garrigan, P., & Kellman, P. J. (2011). The role of constant curvature in 2-D contour shape representations. *Perception, 40,* 1290–1308. http://dx.doi.org/10.1068/p6970

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, Massachusetts: Houghton-Mifflin.

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *The Journal of Neuroscience, 35,* 10005–10014. http://dx.doi.org/10.1523/JNEUROSCI.5023-14.2015

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).

Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception* (pp. 309–331). New York, NY: Holt, Rinehart, & Winston.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195,* 215–243. http://dx.doi.org/10.1113/jphysiol.1968.sp008455

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology, 73,* 218–226. http://dx.doi.org/10.1152/jn.1995.73.1.218

Kellman, P. J., & Arterberry, M. E. (2000). *The cradle of knowledge: Development of perception in infancy*. London, England: MIT press.

Kellman, P. J., Garrigan, P., & Erlikhman, G. (2013). Challenges in understanding visual shape perception and representation: Bridging subsymbolic and symbolic coding. In S. Dickinson & Z. Pizlo (Eds.), *Shape Perception in Human and Computer Vision* (pp. 249–274). London, UK: Springer. http://dx.doi.org/10.1007/978-1-4471-5195-1_18

Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. *Psychology of Learning and Motivation, 58,* 117–165. http://dx.doi.org/10.1016/B978-0-12-407237-4.00004-9

Koffka, K. (1935). *Principles of Gestalt psychology*. New York, NY: Harcourt Brace.

Köhler, W. (1929). *Gestalt psychology*. New York, NY: Liveright.

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science, 1,* 417–446. http://dx.doi.org/10.1146/annurev-vision-082114-035447

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 1097–1105). Red Hook, NY: Curran Associates, Inc.

Landau, B. (1999). Reinventing a broken wheel. *Behavioral and Brain Sciences, 22,* 623–624. http://dx.doi.org/10.1017/S0140525X99372149

Lezama, J. (2015). *On grouping theory in dot patterns, with applications to perception theory and 3D inverse geometry* (Doctoral dissertation). École normale supérieure de Cachan-ENS Cachan, France.

Maunsell, J. H., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology, 68,* 1332–1344. http://dx.doi.org/10.1152/jn.1992.68.4.1332

Michotte, A., Thinés, G., & Crabbè, G. (1964). *Amodal completion and perceptual organization (Tr.)*. Louvain, France: Studia Psychologica.

Ohlsson, S. (1999). Selecting is not abstracting. *Behavioral and Brain Sciences, 22,* 630–631. http://dx.doi.org/10.1017/S0140525X99462144

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology, 86,* 2505–2519. http://dx.doi.org/10.1152/jn.2001.86.5.2505

Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision Research, 37,* 1217–1241. http://dx.doi.org/10.1016/S0042-6989(96)00220-9

Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic Gestalts. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1331–1349. http://dx.doi.org/10.1037/a0024330

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 422–435. http://dx.doi.org/10.1037/0096-1523.3.3.422

Prins, N., & Kingdom, F. A. A. (2009). *Palamedes: Matlab routines for analyzing psychophysical data.* Retrieved from www.palamedestoolbox.org

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387,* 281–284. http://dx.doi.org/10.1038/387281a0

Schultz, D. W., & Eriksen, C. W. (1977). Do noise masks terminate target processing? *Memory & Cognition, 5,* 90–96. http://dx.doi.org/10.3758/BF03209198

Sebastian, T. B., & Kimia, B. B. (2005). Curves vs. skeletons in object recognition. *Signal Processing, 85,* 247–263. http://dx.doi.org/10.1016/j.sigpro.2004.10.016

Sha'ashua, A., & Ullman, S. (1988). Structural saliency: The detection ofglobally salient structures using a locally connected network. In *Proceedings of the 2nd International Conference on Computer Vision* (pp. 321–327). Washington, DC: IEEE Computer Society Press.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Retrieved from arXiv preprint arXiv: 1409.1556

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores? *PLoS ONE, 3*(2), e1699. http://dx.doi.org/10.1371/journal.pone.0001699

Smithson, H., & Mollon, J. (2006). Do masks terminate the icon? *The Quarterly Journal of Experimental Psychology, 59,* 150–160. http://dx.doi.org/10.1080/17470210500269345

Smits, J. T., & Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception, 16,* 121–131. http://dx.doi.org/10.1068/p160121

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74,* 1–29. http://dx.doi.org/10.1037/h0093759

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19–31. http://dx.doi.org/10.1177/001872086300500103

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica, 27,* 285–292. http://dx.doi.org/10.1016/0001-6918(67)90070-4

Subramaniam, S., Biederman, I., & Madigan, S. (2000). Accurate identification but no priming and chance recognition memory for pictures in

RSVP sequences. *Visual Cognition, 7,* 511–535. http://dx.doi.org/10.1080/135062800394630

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. Retrieved from arXiv preprint arXiv:1312.6199

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 2744–2749. http://dx.doi.org/10.1073/pnas.1513198113

Uttal, W. R. (1973). The effect of deviations from linearity on the detection of dotted line patterns. *Vision Research, 13,* 2155–2163. http://dx.doi.org/10.1016/0042-6989(73)90193-4

Wertheimer, M. (1923). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A Source Book of Gestalt Psychology* (pp. 71–88). Gouldsboro, Pennsylvania: The Gestalt Journal Press, Inc.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 8619–8624. http://dx.doi.org/10.1073/pnas.1403112111

Zhang, N. R., & von der Heydt, R. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *The Journal of Neuroscience, 30,* 6482–6496. http://dx.doi.org/10.1523/JNEUROSCI.5168-09.2010

Zhu, Z., Xie, L., & Yuille, A. (2016). Object recognition with and without objects. Retrieved from arvVix preprint arXiv:1611.06596

Appendix A

Method for Generating Shape Contour Stimuli

1. Begin by generating a circle with a radius of two degrees of visual angle in the center of the screen.
2. Select 12 control points along the circle's circumference. Choose control points that are 30 degrees apart along the circle, but jittered in either direction (M = 3.832, SD = 0.6509).
3. For each control point, randomly select an amplitude of displacement from a uniform distribution between 0 and 2.807 degrees of visual angle. Displace control points by the amplitude.
4. Fit cubic splines between the control points and transform from polar to cartesian coordinates.
5. Sample 25 evenly spaced dots from the new shape contour, and color each dot black or white with the constraint that no more than two consecutive dots can have the same color.

Appendix B

Method for Deforming Shape Contours for the "Different" Shape Condition

1. Begin with the shape contour presented in the first display.
2. Calculate the length of the shape contour.
3. Pick one of the 12 control points and displace it a random distance from the center. This distance is sampled from a uniform distribution between 1.289 and 2.740 degrees of visual angle.
4. Pick an adjacent control point and displace it from the center by a distance such that the difference in total contour length between the new shape outline and the original will be minimized.
5. Fit cubic splines between the control points and transform from polar to cartesian coordinates.
6. Sample 25 evenly spaced dots from the new shape contour, and color each dot black or white with the constraint that no more than two consecutive dots can have the same color.

Received June 9, 2017
Revision received December 5, 2017
Accepted January 10, 2018 ■