Research Article

SEEING SETS: Representation by Statistical Properties

Dan Ariely
Massachusetts Institute of Technology

Abstract—*Sets of similar objects are common occurrences—a crowd of people, a bunch of bananas, a copse of trees, a shelf of books, a line of cars. Each item in the set may be distinct, highly visible, and discriminable. But when we look away from the set, what information do we have? The current article starts to address this question by introducing the idea of a set representation. This idea was tested using two new paradigms: mean discrimination and member identification. Three experiments using sets of different-sized spots showed that observers know a set's mean quite accurately but know little about the individual items, except their range. Taken together, these results suggest that the visual system represents the overall statistical, and not individual, properties of sets.*

Sets of similar objects are common occurrences—a crowd of people, a bunch of bananas, a copse of trees, a row of fence posts, a shelf of books, a line of cars. Each item in the set is distinct, highly visible, and discriminable. But when we look away from the set, what information do we have? If something in the set particularly catches our eye, we may retain quite a bit of information about it. A rich representation will also be formed if we had some appropriate categories available—such as the names of several people in the crowd. However, when the stimuli do not fall into preformed categories, what information do we have about each of the individual items? What information do we have about the set of items as a whole? Our language supports the idea that a crowd, a bunch, a copse, or a set is somehow different from the sum of its parts. The work reported here examined whether the human visual system makes similar distinctions and represents properties of items in sets as individual items or as a whole. The experiments begin to address the question: Does the visual system create a specific representation for a set of similar objects that is not just the sum of the representations of the individual items? Two new paradigms were used to determine what observers know about the members of a set and what they know about the statistical properties of the set (mean and distribution).

Comparing the accuracy with which observers represent information about parts and sets allows us not only to answer these general questions, but also to distinguish between two general approaches to understanding the perception and representation of multiple items. The first approach suggests that when many items are presented, the visual system encodes only low-resolution information about each of them (Neisser, 1967). Such an approach is consistent with limited-capacity models of visual processing: If the visual system has a bottleneck at some point, then the more items there are to process at one time, the fewer bits there are available to represent each one (Nakayama, 1990). The second approach suggests that there may be more efficient ways of dealing with limited capacity than simply reducing the resolution of the local representation. The idea here is that sets of objects could be represented in a qualitatively different way than single items. Therefore, when presented with a set of objects, the visual system does not face trade-offs involving the resolution for encoding local information, but rather decides how to divide resources between the two types of representations (individual and set).

GENERAL METHOD

This article reports the results of three experiments that were identical in many regards. In particular, the sets and test stimuli in all the experiments consisted of circular spots of various sizes. Such sets have the advantage that the members do not fall into distinct categories, as they could if they varied in color, shape, or orientation. Two of the experiments measured knowledge about the sizes of the individual spots in a set (member-identification experiments), and one measured sensitivity to the mean size of a set (mean-discrimination experiment). In all the experiments, a set of spots was presented in the first temporal interval of a two-interval trial, and a test stimulus, consisting of one or two test spots, was presented in the second interval. An example of a stimulus pair is shown in Figure 1. Both temporal intervals in a trial were 500 ms in duration. There was no blank time between intervals. The 2 observers were male undergraduate students who had normal vision and were naive as to the purpose of the experiment. The same 2 observers participated in all three experiments. No feedback about the correctness of responses was given in any experiment.

Each set consisted of spots of four sizes that were equally spaced on a log scale. Each size was separated from the next size by a factor of either 1.05 or 1.4 (n). The mean spot diameter was 0.25°. The 1.05 sets included spots of similar sizes—diameters ranged from 0.23° to 0.27°; the 1.4 sets included spots with dissimilar sizes—diameters ranged from 0.15° to 0.42°. These two sets are referred to as sets of similar and dissimilar spots, respectively. Sets with 4, 8, 12, and 16 spots were used, with 1, 2, 3, or 4 spots of each of the four sizes, respectively. The spatial arrangements of the sets were random, with constraints on overall area and minimum proximity between spots to make the average density roughly constant. An example of a set of dissimilar spots (factor of 1.4) is shown in Figure 1 together with a sample test spot. The range of sizes of the test spots exceeded the range of the spots in the sets by the difference between the spot sizes themselves (the factor size).

Each set of spots (and its corresponding test spots) was presented in 15 versions: five differently randomly scaled versions, each presented in three different spatial arrangements. The different versions were used to discourage the observers from basing their judgments on previously seen stimuli. In the data analysis and presentation of results, size of the test spot is represented relative to the mean of the set with which it was presented, averaged over the 15 versions.

Address correspondence to Dan Ariely, Massachusetts Institute of Technology, 38 Memorial Dr., E56-329, Cambridge, MA 02142; e-mail: ariely@mit.edu.

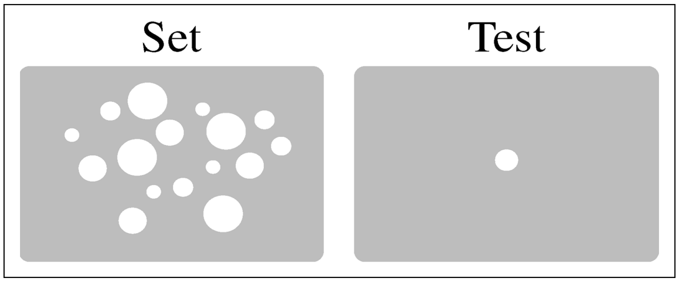

Fig. 1. Schematic representation of the two intervals used in each trial. Observers first saw a set of circles for 500 ms, and then a test stimulus consisting of one or two test spots. This example shows a set of 16 items with a similarity factor of 1.4, and a single test spot.

MEMBER-IDENTIFICATION EXPERIMENTS

Method

In the member-identification experiments, two methods were used: yes/no and two-alternative forced choice (2AFC). In the yes/no experiment, each trial consisted of a presentation of a set of spots in the first interval and a presentation of a single test spot in the second interval (as in Fig. 1). The observer was asked to report whether the test spot was, or was not, a member of the set. Only sets of dissimilar spots were used. For each set size, 150 trials were conducted (10 trials for each of the 15 versions). The percentage of "yes" responses (i.e., the test spot was judged to be a member of the set) was recorded and analyzed.

For each trial in the 2AFC experiment, a set of spots was presented in the first interval (as in the left panel of Fig. 1), and two test spots were presented in the second interval. The size of one of the test spots matched that of one member of the set, and the size of the other test spot was either the next larger or the next smaller in the series of test spots. The observer was asked to report which of the two test spots was a member of the set. Again, only sets of dissimilar spots were used, with set sizes of 4, 8, 12, and 16. For each set size, 150 trials were conducted (10 trials for each of the 15 versions). The percentage of correct responses in the 2AFC task was recorded and analyzed.

Results

Results of the yes/no member-identification experiment are shown in Figure 2. These graphs show the percentage of trials in which the observer responded that the test spot was a member of the set as a function of the size of the test spot. The data were normalized to the maximum percentage of "yes" responses for each observer and set size, making it possible to focus on the distribution of results without regard to the observer's criterion.

The arrows in Figure 2 mark the test spots that were members of

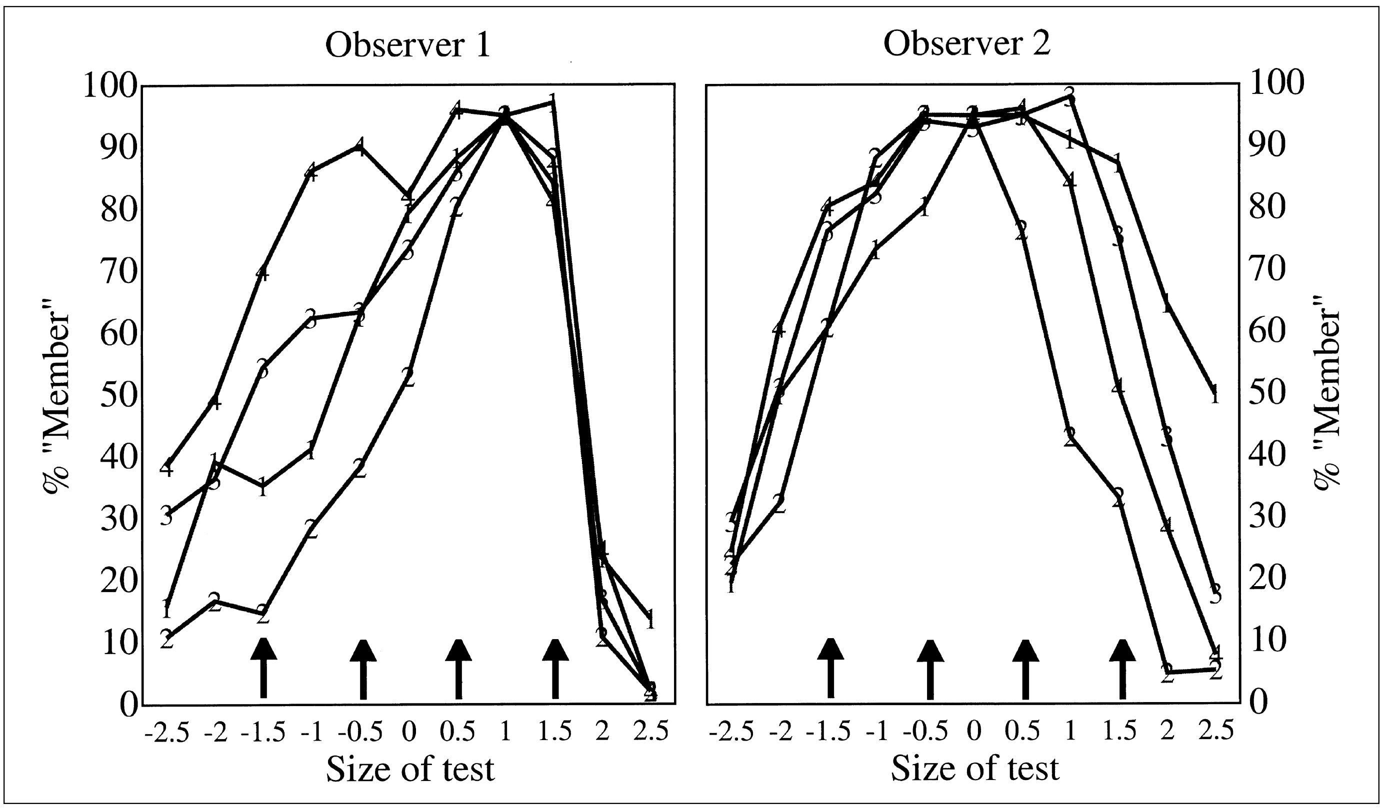

Fig. 2. Yes/no member-identification results. The graphs show the percentage of trials on which the observers identified (correctly or incorrectly) a test spot as a member of the set. Results for the 2 observers are shown separately. Data are shown for four set sizes, with the number on each curve indicating the number of spots of each size (i.e., set size/4). Spot size = $(0.25 \times 1.4^{n})°$. The *x*-axis shows spot size in terms of the value of the exponent (*n*), which is the number of steps to the mean. The arrows mark the test spots that were members of the set; all other test spots were not members of the set.

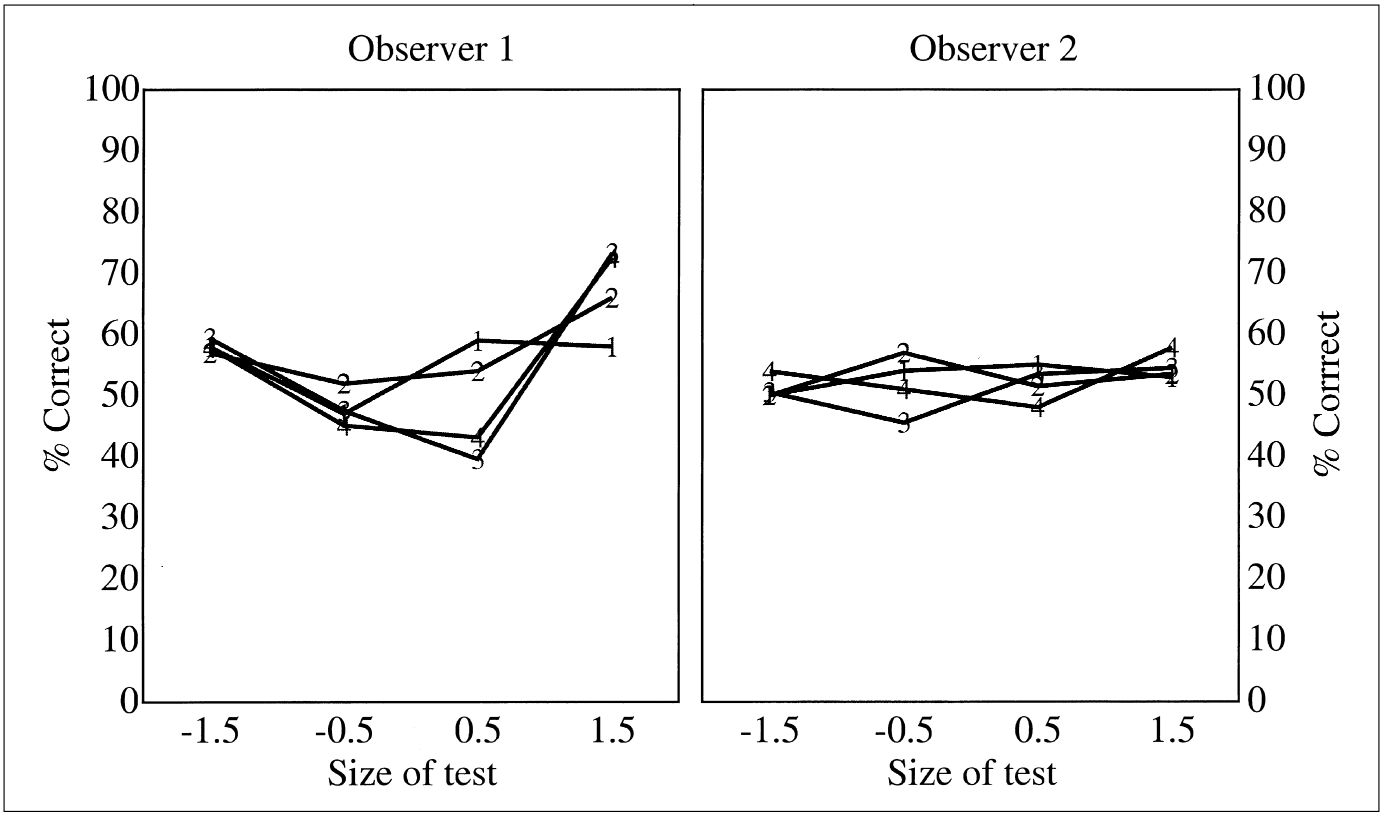

Fig. 3. Results of the two-alternative forced-choice member-identification experiment. The graphs show the percentage of trials on which the observers correctly identified the member in the pair of test items. Data from the 2 observers are shown separately. Data are shown for four set sizes, with the number on each curve indicating the number of spots of each size (i.e., set size/4). Spot size = $(0.25 \times 1.4^n)°$. The *x*-axis shows spot size in terms of the value of the exponent (n), which is the number of steps to the mean.

the set. Perfect performance on this task would have resulted in a sawtooth shape with all data points above the arrows at 100% and all other data points at 0%. The smooth curves obtained differ sharply from that level of performance.

The 2 observers did not differentiate between members and nonmembers within the range of a set. The observers were unable to make this distinction despite the fact that the members differed in size from the nonmembers by at least 18%, about three times the size-discrimination threshold for sets of same-size spots.[1] Moreover, this failure to discriminate held even for the sets with only four items. As can be seen in Figure 2, both observers exhibited some knowledge of the range of the set: They consistently responded “yes” more frequently to test spots that were within the range of the set than to test spots that were outside the range of the set. Within the range of the set (values of the spot size exponent, n, in the range -1.5 to 1.5), the percentage of “yes” responses for most spot sizes was above the 50% line, whereas outside that range, the percentage for most spot sizes was below the 50% line.

The same experiment was repeated with a 2AFC paradigm to provide a criterion-free test of observers’ knowledge about the sizes of the individual spots in the set. In this experiment, each test stimulus consisted of two spots: a member of the set and a nonmember (as described in the Method section). The observers’ task was to report which test spot was a member of the set they had just seen. This is a criterion-free procedure; observers’ subjective estimates of what knowledge they have play no role. Instead, observers are forced to use what information they have to make a judgment. Each member test spot was presented with a larger nonmember test spot half the time and with a smaller nonmember the other half of the time.

The results are shown in Figure 3, which graphs the percentage of times that each member spot was chosen in all of the trials in which it was presented—with a larger spot or with a smaller spot. This averaging across the size of the nonmember test spot eliminates observer bias in the data (e.g., the tendency to choose the spot size that is closer to the mean). The results from the 2AFC procedure show that the observers were unable to distinguish test spots that were in the set from those that were not: Performance of both observers was only marginally better than chance. Overall, it seems that observers are not able to make accurate judgments regarding parts of a set.

MEAN-DISCRIMINATION EXPERIMENT

Method

The stimuli used in the mean-discrimination experiment were identical to those used in the yes/no member-identification experi-

1. A small study was conducted to examine the discrimination threshold of a single spot compared with a set of same-size spots. The results showed discrimination thresholds between 4.5 and 5.8.

ment. Each trial consisted of presentation of a set of spots in the first interval and presentation of a single test spot in the second interval (as in Fig. 1). The observer's task was to report whether the test spot was larger or smaller than the mean spot size of the set. For each of the sets, 330 trials were conducted, including 30 trials for each of 11 test stimuli spanning the likely range of the psychometric function. A standard probit analysis (Finney, 1971) was carried out to determine the mean-discrimination threshold (the standard deviation of the best-fitting cumulative normal distribution).

Results

The mean-discrimination thresholds are shown in Figure 4. In each case, the threshold was either roughly constant across set size or decreased slightly with increasing set size. For the sets with the small, 1.05, step size, the mean-discrimination threshold was quite low—about 4 to 6% of the spot size. Increasing the range dramatically, so that the spots were sized differently enough that they could be considered dissimilar, elevated the threshold to only 6 to 12%. The results show that the mean size of sets was known quite precisely for sets of both step sizes. Further, the results indicate that precision either was independent of the number of items in the set or improved with increasing set size.

Note that an ideal observer who has an error in perception would also produce better mean discrimination than membership discrimination as a result of averaging of errors. Thus, in order to claim that perception of the mean is more accurate than perception of membership, one has to show that the same ideal observer could not produce both sets of results. A simulation of an ideal observer was constructed to allow examination of this issue. In this simulation, an ideal observer perceived each circle with an error that was *k*% of the circle size and made its judgments in the two tasks. The question was whether an ideal observer with a given *k* could produce both sets of results. The analysis showed that a *k* equal to 12 would be needed to explain the results of the mean-discrimination experiment, but the membership-identification experiment would require that *k* equal 40. In other words, although averaging multiple circles can account for a part of the effect, it cannot explain it entirely. Moreover, there are a few other aspects of the results that cannot be explained by the model of an ideal observer. First, such a model would not be able to account for the insensitivity to set size in the member-identification task (see Figs. 2 and 3). Second, it could not account for the constant (and slightly increased) accuracy across set sizes (see Fig. 4). The fact that such errors in perceptions cannot account for the current set of results might be an indication that errors in these types of judgments do not occur during initial perception, but rather occur in later stages.

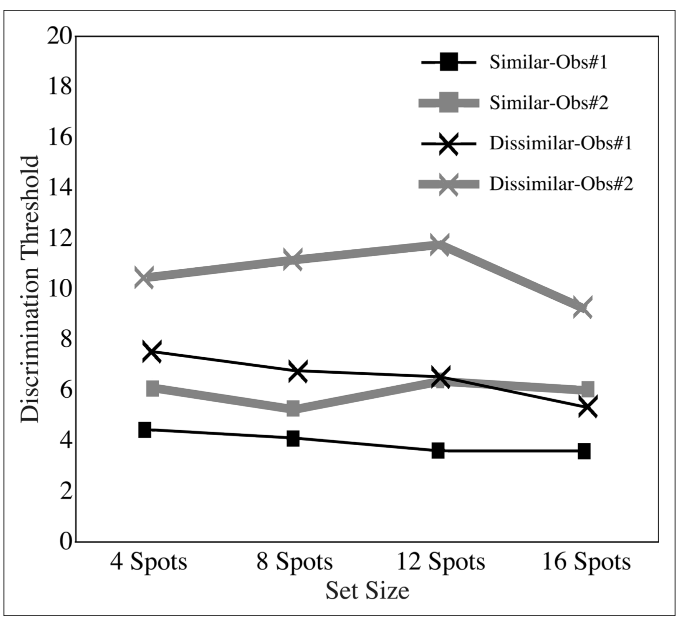

Fig. 4. Mean discrimination thresholds for the four set sizes. Data are shown separately for the 2 observers ("Obs") and the two set distributions (similar and dissimilar).

In sum, the results of the mean-discrimination experiment are surprising, particularly considering the results of the member-identification experiments. If the 2 observers did not have accurate information about the individual items in the set, how is it that they had accurate information about the set? The answer to this question could lie, in part, in the way the visual system represents sets of items. One implication of these results is that the representation of a set is not a simple composition of its discrete parts.

DISCUSSION

Two new experimental paradigms (mean discrimination and member identification) were used to help determine when a given property of a collection of items is represented as a set property and when a specific item is encoded by the visual system as belonging to a set. The results of the member-identification experiments show that observers have a poor ability to discriminate members from nonmembers: Observers knew little or nothing about the sizes of the individual items in a set. In contrast, the results of the mean-discrimination experiment indicate that observers encode quite precise information about the mean of a set. Based on these results, one can speculate that when presented with a set of four or more similar items, the visual system creates a representation of the set, and discards information about the individual items in the set.

Studies in object perception often have as their goal the understanding of how individual items are represented (e.g., Biederman, 1987). The expectation is that the representation of complex scenes will consist of many such individual representations. The present results suggest that the visual system may not take this approach to represent the entire scene. In the many cases in which proximal items are somewhat similar, the representation of a set may instead contain information about, for example, the average size, color, orientation, aspect ratio, and shape of the items in the set and essentially no information about the individual items.

The current work demonstrates the existence of a novel kind of representation: representation of sets of similar items. Such representations do not include information about individual items in a set, but do include highly precise information about the mean of the set and perhaps some information about the range or spread of the set. Orientation, aspect ratio, mean hue, and shape (however represented) may all be set properties—as well as item properties. (Tests of orientation using short line segments showed that observers know the mean orientation of a set of lines, but not individual orientations, a finding consistent with there being a set representation of this property.) Velocity may also be a set property (see also Rosenholtz, 1999). The mean and variance of a set of moving dots are known quite accurately (Atchley & Andersen, 1995; Watamaniuk & Duchon, 1992); it is reasonable to guess that the velocities of the individual dots are not. I turn now to a discussion of some other implications of such set representations.

The Economy of Set Representations

Although the exquisite sensitivity of the human visual system to minute differences in sensory input is a source of wonder and delight to us, a computer can readily be equally sensitive. Much harder to solve, and currently far beyond the abilities of computer vision systems, is the problem of knowing what information to throw away. We know that contrast is more important than luminance (Mach, 1914/1959), and that ratios and differences—not quantal counts—determine color (Hering, 1964). Relationships matter more than absolutes (Helson, 1964). But in research on spatial vision, the focus has been more on reproducing the image exactly, on compressing the representation without losing information (e.g., Watson, 1987). Although this focus is useful in practical applications, it does not answer the basic question: What spatial information is actually represented and in what form? If our models do not lose any of the original data, then no decisions have been made; no information has been extracted (Julesz, 1995). The reduction of a set of similar items to a mean (or prototypical value), a range, and a few other important statistical properties may preserve just the information needed to navigate in the real world, to form a stable global percept, and to identify candidate locations of interest.

Visual Search as Set Segregation

Closely related to the study of set representation is the research on visual search: the investigation of what determines the speed with which a target item can be located when embedded in many other items, called the distractors. The central—and initially surprising—finding in this field is that a target that is readily discriminable from a single distractor may not be so discriminable (i.e., a relatively long time may be needed to locate it) when it is embedded in multiple copies of that distractor.

This key result can be explained in terms of the concept of set representation. Distractors (and sometimes the target, too) in a visual search task might be represented as set properties (i.e., by their mean, variance, and other set properties). One can further speculate that what is known as preattentive, parallel, or set-size-independent search occurs when the representation of the target is sufficiently different from the set representation. This basic notion can suggest new interpretations for several basic visual search results. The implication for increased variance in the distractors (Duncan & Humphreys, 1989) is obvious. The variance of the set representation is larger when there is more variability in the distractors themselves (provided that variability exceeds the intrinsic variance). Consequently, as the variance of the set increases, a target must differ more from the mean to segregate from the set (see also Rosenholtz, 1999).

The considerable research on conjunctive searches (search for a target that differs from the distractors only in its combination of features) also takes on a different look when thought of in terms of sets. A set representation is associated with the full spatial region of the set, not with individual items, because the individual items are not represented. Thus, for example, a set of red and blue items has both colors attributed to the entire set. A violet item would be a member of this set—assuming it is encoded as a red-and-blue item—and consequently would not be readily locatable—and indeed, it is not (Treisman, 1991). Similarly, the difficulty of finding a target that is defined by the absence of a feature, such as finding an *O* among *Q*s, is explainable in terms of a set representation. Again, all one needs is the assumption that set representations are associated with the full spatial region of the set (see also Grossberg, Mingolla, & Ross, 1994; Wolfe, Cave, & Franzel, 1989).

In addition to posing intriguing questions of their own, visual search studies have been useful in ferreting out the dimensionality of the underlying representations. The mean-discrimination and member-identification paradigms presented here provide a more direct approach to this problem. These paradigms allow us to directly measure the groupings the visual system makes, so that we do not have to make inferences from the results of a search task.

Visual Contrast Illusions and Set Representations

An additional related area to which set representation can be applied is contrast and grouping. Consider, for example, the well-known Ebbinghaus illusion (see Fig. 5, top panel). It has been shown that the central circle on the left seems smaller than the central circle on the right. Examining this phenomenon from the perspective of set representation suggests that the central circle on the left is judged relative to the set properties of the circles surrounding it, whereas the central circle on the right is judged relative to the set properties of the circles surrounding it. If the basic representation of a set contains the relationship of an object to the mean of its set, it follows that the central circle on the left will be judged smaller than the central circle on the right. Moreover, if the central circle were not perceived as belonging to the same set as the sur-

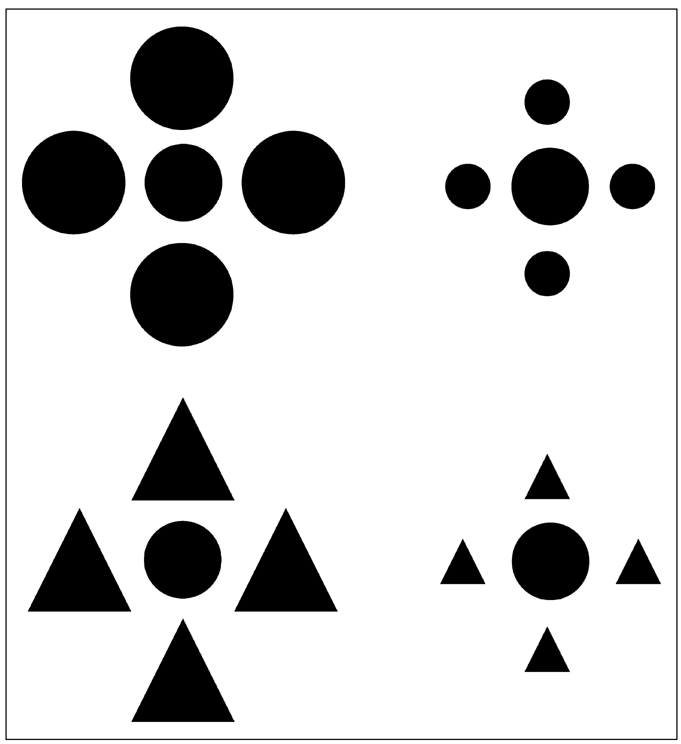

Fig. 5. The Ebbinghaus illusion and its attenuation. In this illusion (top panel), the central circle in the array on the left is seen as smaller than the central circle in the array on the right. The illusion is attenuated when the surrounding circles are changed to triangles (bottom panel).

rounding objects, the Ebbinghaus illusion would be attenuated. Coren and Miller (1974) obtained exactly this result (see Fig. 5, bottom panel).

Statistical Representation of Experiences

The idea of representation by statistical properties may be even more broadly useful. Consider, for example, nonvisual experiences such as a dental treatment, a meal, and a college course. A feature of such experiences is that they unfold over time through a stream of transient states that may vary in intensity and even in sign from moment to moment. Just as we might ask how people represent sets of similar items, we can ask how people represent such experiences that are composed of multiple parts. In particular, we can ask whether such experiences are represented by their individual components, or whether they are represented by a few overall statistical properties.

Recent work (Kahneman, 2000) has demonstrated that when people summarize experiences, they do not simply combine the intensities of the actual experiences. Rather, two kinds of defining features (gestalt characteristics) appear to be given particularly high weight (importance). One reflects the change over time in the intensity of the transient states. Prominent examples of such characteristics include the trend of the profile (Ariely, 1998) and its rate of change (Hsee & Abelson, 1991). The other type of gestalt characteristic reflects the intensity of the transient experience at particular key points in time. Specifically, a variety of studies have found that the momentary experience at the most intense and final moments (peak and end, respectively) can account for global retrospective evaluations (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). This work on hedonic calculus (how people combine experiences of pleasure and pain) suggests that the set representations proposed here may not be limited to the visual system. Other types of experiences, such as touch, taste, sounds, and even pleasure and pain, may be represented by some higher-level statistical properties.

Acknowledgments—I would like to express my deep gratitude to Christina Burbeck for her help and guidance, Gal Zauberman for stimulating discussions and help with the experiments, and Gregory Ashby for his suggestions about the role of error in perceptual judgments.

REFERENCES

Ariely, D. (1998). Combining experiences over time: The effects of duration, intensity changes and on-line measurements on retrospective pain evaluations. *Journal of Behavioral Decision Making*, *11*, 19–45.
Atchley, P., & Andersen, G. (1995). Discrimination of speed distributions: Sensitivity to statistical properties. *Vision Research*, *35*, 3131–3144.
Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychology Review*, *94*, 115–147.
Coren, S., & Miller, J. (1974). Size contrast as a function of figural similarity. *Perception & Psychophysics*, *16*, 355–357.
Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.
Finney, D.J. (1971). *Probit analysis*. Cambridge, England: Cambridge University Press.
Grossberg, S., Mingolla, E., & Ross, W.D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface and spatial object representations. *Psychological Review*, *101*, 470–489.
Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper and Row.
Hering, E. (1964). *Outlines of a theory of the light sense* (L.M. Hurvich & D. Jameson, Trans.). Cambridge, MA: Harvard University Press.
Hsee, C.K., & Abelson, R.P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality and Social Psychology*, *60*, 341–347.
Julesz, B. (1995). *Dialogues on perception*. Cambridge, MA: MIT Press.
Kahneman, D. (2000). Evaluation by moments: Past and future. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames* (pp. 699–708). New York: Cambridge University Press and the Russell Sage Foundation.
Kahneman, D., Fredrickson, B.L., Schreiber, C.A., & Redelmeier, D.A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*, 401–405.
Mach, E. (1959). *The analysis of sensation.* New York: Dover. (Original work published 1914)
Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 135–149). Cambridge, England: Cambridge University Press.
Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.
Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, *39*, 3157–3163.
Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 652–676.
Watamaniuk, S.N.J., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*, 931–942.
Watson, A.B. (1987). The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics, Image Processing*, *39*, 311–327.
Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.

(RECEIVED 12/8/99; REVISION ACCEPTED 7/25/00)