

## Attention to distinguishing features in object recognition

Orit Baruch, Ruth Kimchi, and Morris Goldsmith

Department of Psychology and Institute of Information Processing and Decision Making, University of Haifa, Haifa, Israel

(Received 18 March 2013; accepted 10 November 2014)

This study advances the hypothesis that, in the course of object recognition, attention is directed to *distinguishing features*: visual information that is diagnostic of object identity in a specific context. In five experiments, observers performed an object categorization task involving drawings of fish (Experiments 1–4) and photographs of natural sea animals (Experiment 5). Allocation of attention to distinguishing and non-distinguishing features was examined using primed-matching (Experiment 1) and visual probe (Experiments 2, 4, 5) methods, and manipulated by spatial precuing (Experiment 3). Converging results indicated that in performing the object categorization task, attention was allocated to the distinguishing features in a context-dependent manner, and that such allocation facilitated performance. Based on the view that object recognition, like categorization, is essentially a process of discrimination between probable alternatives, the implications of the findings for the role of attention to distinguishing features in object recognition are discussed.

**Keywords:** Object recognition; Object categorization; Visual attention; Distinguishing features.

The role of visual attention in object recognition has been a subject of debate. Early-selection models of attention (Broadbent, 1958; Lachter, Forster, & Ruthruff, 2004) hold that unattended objects are not processed beyond the analysis of simple physical features. Along these lines, Feature Integration

---

Please address all correspondence to Orit Baruch, Department of Psychology and Institute of Information Processing and Decision Making, University of Haifa, Haifa 31905, Israel. E-mail: [oritb@research.haifa.ac.il](mailto:oritb@research.haifa.ac.il)

This research was supported by a grant [grant number 1110/10] from the Israel Science Foundation to Ruth Kimchi and Morris Goldsmith, and comprises part of Orit Baruch's PhD dissertation, submitted to the University of Haifa. Facilities for conducting the research were provided by the Institute of Information Processing and Decision Making and by the Max Wertheimer Minerva Center for Cognitive Processes and Human Performance.

Theory (Treisman & Gelade, 1980) proposed that attention is crucial for the perception of an integrated object, as it operates to bind featural information represented in independent feature maps. More recently, the Reverse Hierarchy Theory (Hochstein & Ahissar, 2002) holds that whereas spread attention enables the initial perception of coherent conjoined objects “at a glance,” based on bottom-up processing alone, top-down focused attention must subsequently be invoked to consciously identify lower-level details such as orientation, colour, and precise location.

In contrast, other studies suggest that object recognition can occur without attention or awareness (e.g., Li, VanRullen, Koch, & Perona, 2002; Luck, Vogel, & Shapiro, 1996). For example, Li et al. (2002) found that rapid detection of animals or vehicles in briefly presented natural scenes was unaffected by an attention-demanding task performed simultaneously. These findings, however, were later found not to generalize to more complex visual scenes that include multiple foreground objects (Walker, Stafford, & Davis, 2008). Moreover, while the mere detection of an animal in a natural scene may require little attention, precise identification of the animal appears to be attention demanding (Evans & Treisman, 2005). Thus, an increasing number of findings demonstrate the need for attention in perceptual tasks that were previously considered to be attention free.

In this work, we advance the idea that attention plays a special role in object recognition, namely, that the allocation of attention to distinguishing features serves to facilitate the extraction of diagnostic object information. The initial information extracted from the visual scene in a data-driven manner (without attention) is inherently limited. In natural scenes, portions of objects are often hidden from view and surfaces undergo occlusion; sometimes the viewing conditions are poor, and at other times the relevant information is subtle and cannot be acquired at a glance. Moreover, even under optimal viewing conditions, the initial information may be partial (e.g., coarse information carried by low spatial frequencies; Bar, 2003; Fabre-Thorpe, 2011; Hughes, Nozawa, & Kitterle, 1996). Although, depending on context, the initial partial information may sometimes suffice for the task at hand (e.g., Johnson & Olshausen, 2005), in many cases the information that can be extracted in such a manner is inadequate.

The notion of features (and hence, distinguishing features) as conceived here is very broad, and refers to any aspect of an object that can serve to discriminate between the set of probable alternatives in a particular context or view. Such aspects may include structural or configural features (e.g., *geons*; Biederman, 1987), surface features (e.g., colour or texture), global features (e.g., global shape: elongated vs. round), or localized features and parts (e.g., the shape or colour of a beak), which may be either view dependent or view invariant. Thus, despite the historical conflict (e.g., Peissig & Tarr, 2007), we do not see attention to distinguishing features as necessarily conflicting with view-based or structural-description based accounts of object recognition.

The idea that distinguishing features provide the basis for recognition and categorization, and that attention is directed to these features, has a long history. Eleanor Gibson (1969) proposed that perceptual learning occurs through the discovery of features that distinguish one pattern from another, and that part of the optimization of attention that occurs with development involves selective attention to these features. Tversky (1977) proposed a “diagnosticity principle”, by which participants attend to the features of a stimulus that have classificatory significance. Several studies have shown that participants are in fact differentially sensitive to diagnostic dimensions (e.g., Aha & Goldstone, 1990; Goldstone, 1994; Nosofsky, 1986), giving these more weight when making categorization judgements. These studies, however, did not include direct measures of attentional involvement.

Indeed, the specific claim that attention is directed to distinguishing features in object recognition has been addressed in relatively few studies. Most of these have used eye tracking as an indirect measure of attention. For example, Rehder and Hoffman (2005a, 2005b; see also Blair, Watson, Walshe, & Maj, 2009; Ganis, Schendan, & Kosslyn, 2007) found that during visual category learning, diagnostic features were fixated more often than non-diagnostic features and that the proportion of correct responses correlated with the time diagnostic features were fixated.

Results showing increased eye fixations to diagnostic visual information during visual category learning can be taken to suggest that attention may be directed to distinguishing features during the process of object recognition. However, overt eye movements are only one, somewhat crude, indicator of visual attention and, in fact, in the studies just mentioned, in order to facilitate the analysis of eye fixations, the artificial visual stimuli were generally large and complex, with categorization decisions taking on the order of several seconds of scrutiny to complete.

To our knowledge, only one study has used a more direct measure of spatial attention in examining attention to distinguishing features—a study that was conducted in the context of word and letter recognition (Navon & Margalit, 1983). Using a visual-probe method, Navon and Margalit found that probe detection rate was highest when the probe appeared near the feature that distinguished between two competing word or letter alternatives.

In the present article, we report a series of five experiments using a set of converging methods (primed matching, visual probe, and spatial cueing) to investigate the allocation of visual attention to distinguishing features in object recognition. As can be seen in the preceding brief review, much of the relevant empirical research until now has used object categorization tasks, in which the objects to be recognized are specific instances of just a few possible objects. Restricting the stimulus set in this way has a clear methodological advantage, in allowing a very high degree of experimental control over the relevant distinguishing (and non-distinguishing) features. Primarily for this reason, the

present study was also designed, at the operational level, using object categorization tasks. Nonetheless, based on the view that object recognition, like categorization, is essentially a process of discrimination between probable alternatives, we believe (and will later argue) that the implications of the findings have a direct bearing on the role of attention to distinguishing features in the process of object recognition generally (see General Discussion). In the meantime, to simplify the terminology and avoid confusion, we use the term “categorization” when describing the specific experimental tasks, but use the term “recognition” when formulating hypotheses and reporting and interpreting results.

## OVERVIEW OF THE EXPERIMENTS

The objects for the categorization task in Experiments 1–4 were two-dimensional line drawings of fish, adopted from Sigala and Logothetis (2002). As just explained, using line drawings of stimulus objects in an object categorization task allows maximal experimental control over the object features (both distinguishing and non-distinguishing), and hence enables precise predictions regarding the dynamics of attention. The fish varied in four basic features: Mouth (M), Tail (T), Dorsal Fin (DF) and Ventral Fin (VF), each having three possible values (see Figure 1). Some of the experiments included additional features such as texture or colour. In each experiment, a subset of these features was used to create different categories of fish. Category names were chosen to sound like real fish names. Thus, for example, all fish having a particular DF might be labelled “Grout,” whereas fish with a different DF would be labelled “Tass.” To prevent an association between local fish features and specific locations on the screen, the fish were presented equally often facing left or right.

In Experiment 5 we attempted to generalize the results to more natural objects, with larger variability in visual information, both between and within categories, creating the “gold standard” problem of object recognition—to identify a specific visual stimulus as an instance of a particular type of object,



**Figure 1.** Examples of fish stimuli used in Experiments 1–4. The fish have four basic feature variables: Mouth (M), Tail (T), Dorsal Fin (DF) and Ventral Fin (VF), with three values each.

despite differences in viewpoint (the problem of viewpoint invariance, e.g., Peissig & Tarr, 2007).

All experiments included a training phase before each categorization task, whose purpose was to acquaint the participant with the relevant object categories. On each trial of the training phase, a single object was presented at the centre of the screen. Participants were asked to recognize the object (e.g., identify it as being a “Grout” or as being a “Tass”) as rapidly and accurately as possible, by pressing the appropriate key. A short, low, auditory feedback was provided after erroneous responses. The training phase ended when the participant made no more than two errors in 20 consecutive trials (i.e., 90% accuracy).

The subsequent categorization task was similar to the training task, but also included other components specific to each experiment. Across the experiments, several different methods were used, including primed matching, visual probe, and spatial cueing, to reveal information about the allocation of attention to distinguishing features during the object categorization-recognition process. Participants were undergraduate students, all with normal, or corrected-to-normal vision. No participant was included in more than one experiment.

## ATTENTION IS DIRECTED TO DISTINGUISHING FEATURES IN THE COURSE OF OBJECT RECOGNITION

This series of three experiments used converging methods to examine the basic hypothesis that in the process of object recognition, attention is directed to distinguishing features—those features that discriminate between probable alternatives. To test this hypothesis, we capitalized on two well-studied phenomena: (1) *Priming* (Experiment 1) refers to the phenomenon by which a recently attended stimulus facilitates performance on a subsequent presentation of that stimulus (e.g., Beller, 1971; Maljkovic & Nakayama, 1994; Tipper, 1985). Thus, if attention is allocated to distinguishing features in the course of object recognition, priming by the object should facilitate responses to stimuli that are similar to its distinguishing features relative to stimuli similar to the non-distinguishing features. (2) *Attentional benefit* (Experiment 2) refers to the phenomenon by which attending to a specific location improves the processing of stimuli presented at that location (e.g., Posner, 1980). Thus, if in the course of object recognition attention is allocated to distinguishing features, the detection of a visual probe should be better when the probe is presented near the distinguishing feature than near a non-distinguishing feature. Also, if attending to distinguishing features in fact aids the recognition process, then advanced allocation of attention to the location of a distinguishing feature (Experiment 3) should yield faster and/or more accurate recognition than if attention is

pre-allocated to the location of a non-distinguishing feature. In this series of experiments the objects were two dimensional line drawings of fish.

## Experiment 1

This experiment examined the allocation of attention to distinguishing features (local features of the contour in Experiment 1a, and surface features in Experiment 1b), using a version of the primed-matching method (Beller, 1971). In this method, a priming stimulus is followed by a pair of test stimuli to match for identity. Responses to correctly identified “same” pairs depend on the representational similarity of the test stimuli to the prime, such that responses to same pairs are relatively faster and/or more accurate when the test stimuli are similar to the prime than when they are dissimilar to it (e.g., Beller, 1971; Rosch, 1975). In the present experiment, on each trial the participant was presented with a single fish to categorize. Immediately following the participant’s categorization response, a pair of test stimuli was presented. There were two types of “same”-response test pairs, one in which the stimuli were similar to the distinguishing feature of the immediately preceding fish and the other in which the stimuli were similar to the non-distinguishing feature of that fish. Note that test pairs that are similar to the non-distinguishing feature are actually dissimilar to the distinguishing feature. Because both types of test pairs are similar to the fish, no difference between responses to these two types of pairs is expected, unless there is an advantage for one feature over the other as when one of the features is attended to. Thus, we reasoned that if attention is allocated to the distinguishing feature in the course of recognizing (categorizing) the fish, priming by a specific fish should facilitate “same” responses for test stimuli similar to its distinguishing feature relative to test stimuli similar to its non-distinguishing feature.

### Experiment 1a

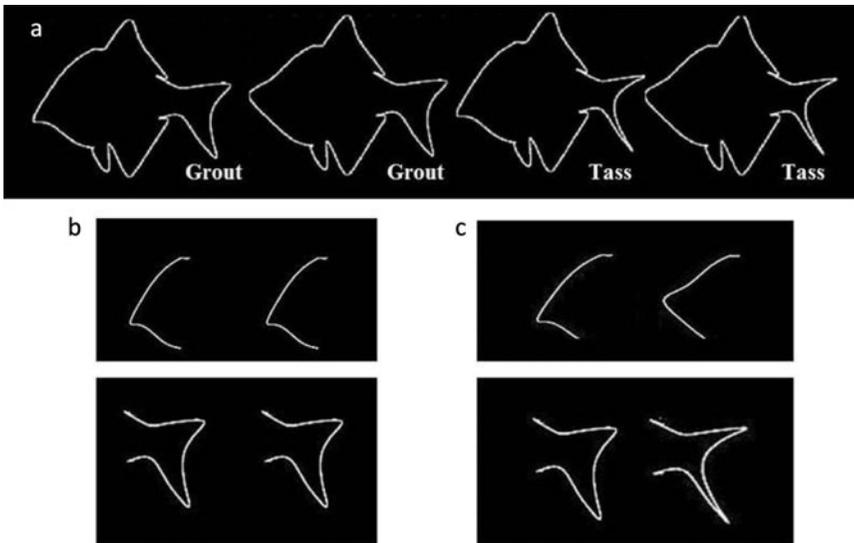
This experiment examines the allocation of attention to local distinguishing features. In one condition the distinguishing feature was the shape of the mouth and in the other condition the shape of the tail. On each trial, participants performed the fish categorization task. Following the categorization response, a test pair was presented to match for identity. The “same”-response test pairs consisted of two outlined mouths or two outlined tails, similar to the mouth or the tail of the immediately preceding fish.

#### *Method*

*Participants.* Twelve students at the University of Haifa participated in the experiment. All had normal or corrected-to-normal vision.

*Stimuli.* A set of four outlined fish was used for the categorization task. The fish varied in the shape of the mouth and the shape of the tail (Figure 2a). In one condition the distinguishing feature was the shape of the tail (tail condition, depicted in Figure 2a), and in the other condition the distinguishing feature was the shape of the mouth (mouth condition). The fish outline was white on a black background. In this experiment and in all subsequent experiments, the maximal span of the fish both in width and height was  $10^\circ$ . The stimuli for the same-different task were outlined mouths and outlined tails (Figures 2b and 2c). The “same”-response test pairs consisted of two identical outlined mouths or two identical outlined tails (Figure 2b). The “different”-response test pairs consisted of two different mouths or two different tails (Figure 2c).

*Design and procedure.* The experiment employed a factorial combination of three factors in a within subject design: distinguishing feature (mouth, tail), test pair (mouth, tail), and response (same, different). The two distinguishing feature conditions were administered in separate blocks. A training phase (described



**Figure 2.** Examples of the stimuli used in Experiment 1a. (a) The set of fish for the categorization task. The fish vary in tail and mouth. The example depicts “Grout” and “Tass” that are distinguished by their tail. (b) Examples of “same”-response test pairs (a pair of two identical mouths and a pair of two identical tails) and (c) “different”-response test pairs, used in the same-different task. The “same”-response pair can be similar to the distinguishing feature or to the non-distinguishing feature of the preceding fish. For example, both “same”-response pairs can follow the left Grout shown in panel a, because both are similar to a feature of this Grout (either the tail or the mouth), and therefore both can be equally primed by the Grout. However, if the distinguishing feature is attended, then priming by this Grout should facilitate “same” responses for test pair similar to its distinguishing feature—the pair of tails in this case, relative to test pair similar to its non-distinguishing feature—the pair of mouths in this case.

above) was administered at the beginning of each block. Each block included 384 trials, preceded by 30 practice trials, with each combination of test pair (mouth and tail) and response (same and different) occurring on an equal number of trials and randomized within block. The order of the distinguishing features was counterbalanced across participants.

On each experimental trial the participants performed two successive tasks: a categorization task followed by a same-different task. Each trial began with the appearance of a small fixation cross for 600 ms, followed by a 400 ms interval. Then a fish was presented at the centre of the screen, facing left or right equally often. The fish remained on the screen until the participant responded. The participant was required to decide as rapidly and accurately as possible whether the fish was a “Grout” or a “Tass” by pressing one of two keys using the left hand. Immediately following the categorization response, a pair of test figures was presented and participants had to judge, as rapidly and accurately as possible, whether the two test figures were the same as each other or different from one another. The test figures remained on the screen until the participant responded by pressing one of two keys using the right hand. The figures in the “same”-response test pairs were similar either to the distinguishing feature (pairs of tails in the tail condition and pairs of mouths in the mouth condition) or to the non-distinguishing feature (pairs of tails in the mouth condition and pairs of mouths in the tail condition) of the immediately preceding fish. Feedback about incorrect response was provided by a short tone.

### *Results and discussion*

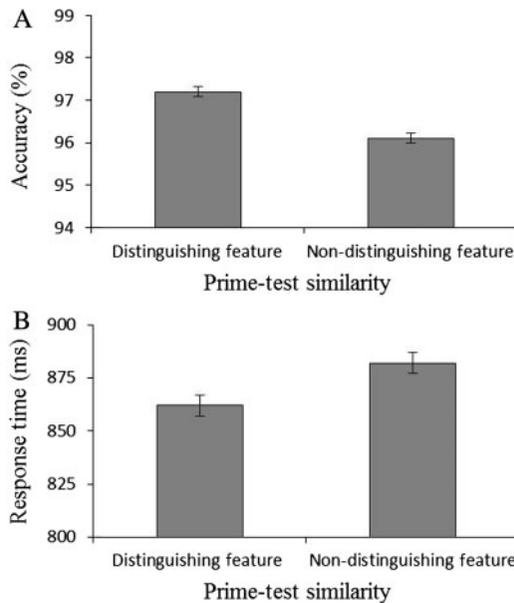
*Categorization task.* Performance in the categorization task was highly accurate (mean accuracy = 96.55%). Participants correctly recognized (categorized) the fish on 96.5% of the trials in the tail condition and on 96.6% of the trials in the mouth condition. The mean response time (RT) was 746 ms and 771 ms, for the tail and mouth conditions, respectively.

*Same-different task.* Accuracy of “same” responses and mean RT for correct “same” responses were calculated for each participant. Only responses that followed a correct categorization response were considered. RTs shorter than 200 ms and longer than 2000 ms (2.4% of all trials) were trimmed.

Computationally, priming is defined as the difference in RT and/or accuracy to respond “same” to test pairs similar to the prime’s (fish) distinguishing feature versus test pairs dissimilar to the distinguishing feature (i.e., similar to the non-distinguishing feature), minus baseline differences between these test pairs. Since in our experiment the prime (fish) was the same in the two distinguishing feature conditions (mouth condition, tail condition) and each of the test pairs (mouth pair, tail pair) was similar to the distinguishing feature of the fish in one condition and similar to the non-distinguishing feature in the other condition, the

response difference between the two prime-test similarity conditions “similarity to the distinguishing feature” and “similarity to the non-distinguishing feature”, collapsed across the two distinguishing features, provides a measure of priming that takes into account any baseline response differences between the mouth and tail test pairs. Faster and/or more accurate “same” responses to test pairs similar to the distinguishing feature than to test pairs similar to the non-distinguishing feature indicate priming of the distinguishing feature; faster and/or more accurate “same” responses to test pairs similar to the non-distinguishing feature than to test pairs similar to the distinguishing feature indicate priming of the non-distinguishing feature.

Mean RT and mean accuracy for each type of “same” response test pair as a function of prime-test similarity (similarity to the distinguishing feature and similarity to the non-distinguishing feature) are presented in Figure 3. Comparisons between the two prime-test similarity conditions were conducted using paired *t*-tests (two-tailed). The analysis showed a significant difference in RT: “same” responses to the test pairs similar to the distinguishing feature were significantly faster (by 21 ms) than “same” responses to the test pairs similar to the non-distinguishing feature,  $t(11) = 2.91$ ,  $p < .014$ ,  $\eta_p^2 = .44$ . The comparison conducted on the arcsine-transformed accuracy data showed similar results for



**Figure 3.** Results from Experiment 1a. (a) Mean accuracy and (b) mean RT for “same”-response test pairs as a function of prime-test similarity: test pairs are either identical to the distinguishing feature or to the non-distinguishing feature of the preceding fish.

accuracy: “same” responses to the test pairs similar to the distinguishing feature were significantly more accurate (by 3.8%) than “same” responses to test pairs similar to the non-distinguishing feature,  $t(11) = 2.87$ ,  $p < .015$ ,  $\eta_p^2 = .43$ .

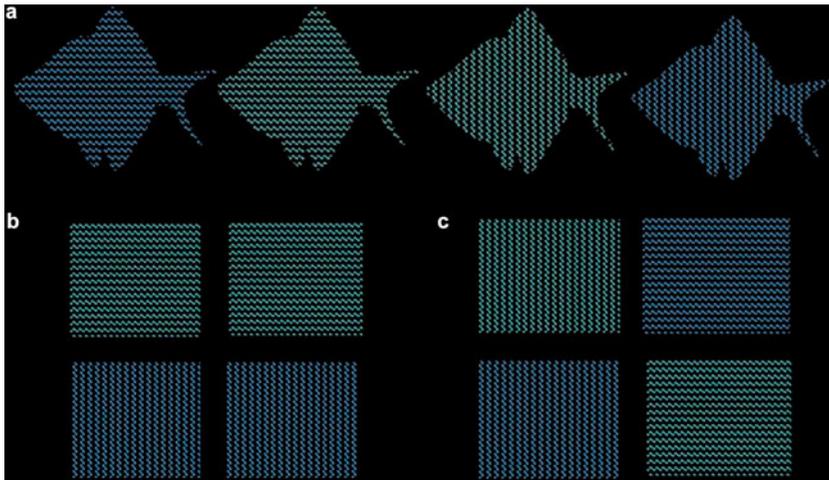
Thus, priming by a fish facilitated “same” responses to test pairs similar to its distinguishing feature relative to test pairs similar to its non-distinguishing feature, indicating that the distinguishing feature was primed. This result suggests that attention was allocated to the distinguishing features during the process of fish recognition.

## Experiment 1b

This experiment examines whether attention is also allocated to distinguishing features that apply to the surface of the object—colour and texture—using a similar procedure as in Experiment 1a.

### *Method*

*Participants.* Ten students at the University of Haifa participated in the experiment. All had normal or corrected-to-normal vision.



**Figure 4.** Examples of stimuli used in Experiment 1b. (a) The set of fish for the categorization task. The fish vary in colour (light or dark blue) and texture (horizontal or vertical stripes). (b) Examples of the “same”-response test pairs used in the same-different task. Both can follow the leftmost fish from panel a: The top pair has the same texture as the texture of this fish and different colour (“texture-similarity” test pair), and the bottom pair has the same colour as this fish and different texture (“colour-similarity” test pair). Priming emerges as a result of attending one of these features when the fish is displayed (i.e., the distinguishing feature). (c) “different”-response test pairs used in the same-different task.

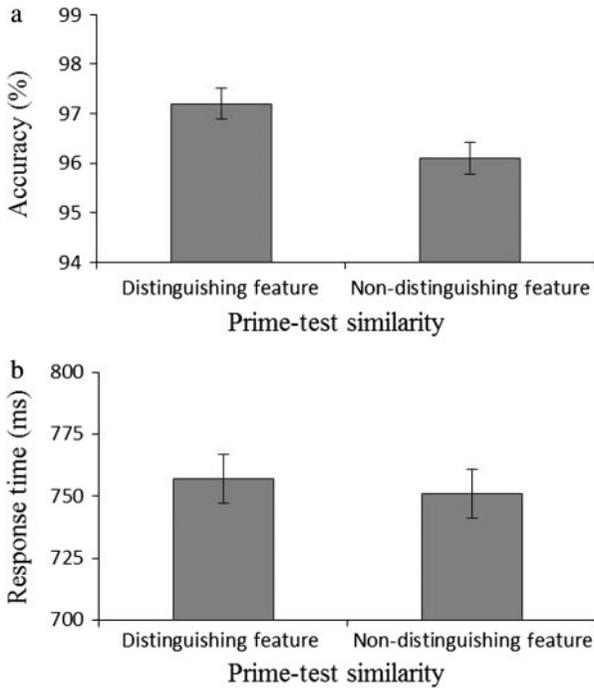
*Stimuli.* The set of fish to be recognized varied in colour and texture (Figure 4a). The texture comprised of local small slanted elements that yielded either horizontal or vertical stripes. The colours were light blue (RGB 90, 155, 160) and dark blue (RGB 70, 120, 160). The stimuli were displayed on a black background. In one condition the distinguishing feature was the colour of the fish (colour condition), and in the other condition the distinguishing feature was the texture (texture condition). The stimuli for the same-different task were rectangles that varied in colour and texture (Figures 4b and 4c). The figures in the “same” response test pairs were identical to one another, but their similarity relation to the immediately preceding fish varied, yielding two types of “same”-response test pairs: a “colour-similarity” test pair in which the colour of the rectangles was similar to the colour of the immediately preceding fish (but the texture was dissimilar), and a “texture-similarity” test pair in which the texture of rectangles was similar to the texture of the immediately preceding fish (but the colour was dissimilar). Thus, a “same” response test pair was similar either to the distinguishing feature (i.e., a “colour-similarity” test pair in the colour condition and a “texture-similarity” test pair in the texture condition) or to the non-distinguishing feature (i.e., a “texture-similarity” test pair in the colour condition and a “colour-similarity” test pair in the texture condition).

*Design and procedure.* The experiment employed a factorial combination of three factors in a within subject design: distinguishing feature (colour, texture), test pair (colour-similarity, texture-similarity), and response (same, different). The two distinguishing feature conditions—the colour condition and the texture condition—were administered in separate blocks and their order was counter balanced across participants. Each block included 400 trials, preceded by 30 practice trials, with each combination of test pair and response occurring on an equal number of trials and randomized within block. All other aspects of the design and procedure were the same as in Experiment 1a.

### *Results and discussion*

*Categorization task.* Performance in the categorization task was highly accurate (mean accuracy = 96.93%). Participants correctly recognized the fish on 97.3% of the trials in the texture condition and on 96.6% of the trials in the colour condition. Recognition was faster in the colour condition (mean RT = 708 ms) than in the texture condition (mean RT = 793 ms),  $F(1,9) = 11.86$ ,  $MSE = 3072.0$ ,  $p < .01$ ,  $\eta_p^2 = .57$ .

*Same-different task.* Accuracy and mean RT for correct “same” responses that followed a correct categorization response were calculated for each participant. RTs longer than 2000 ms and shorter than 200 ms were trimmed (1.2% of all trials). As in Experiment 1a, priming was defined as the difference in RT and/or accuracy between “same” responses to test pair similar to the distinguishing



**Figure 5.** Results from Experiment 1b. (a) Mean accuracy and (b) mean RT for “same”-response test pairs as a function of prime-test similarity.

feature versus test pair similar to the non-distinguishing feature, collapsed across the two distinguishing features. Comparisons between the two prime-test similarity conditions were conducted using paired  $t$ -tests (two-tailed). A significant priming effect occurred only in accuracy. Mean accuracy and mean RT for the two prime-test similarity conditions are presented in Figure 5. A comparison between the two conditions conducted on the arcsine transformed accuracy data showed that “same” responses for test pairs similar to the distinguishing feature were significantly more accurate (by 1.2%) than “same” responses to test-pairs similar to the non-distinguishing feature,  $t(9) = 3.14$ ,  $p < .012$ ,  $\eta_p^2 = .52$ , indicating priming of the distinguishing feature. No significant difference in RT was found: “same” responses to the test pairs identical to the non-distinguishing feature were faster (by 6 ms) than “same” responses to the test pairs identical to the distinguishing feature,  $t(9) < 1$ .

In contrast to Experiment 1a, in which the priming effects were observed both in RT and in accuracy, priming in the present experiment was evident only in accuracy, suggesting a somewhat weaker effect. This may be due to the difference in the nature of the distinguishing features between the two experiments. In Experiment 1a the distinguishing feature and the

non-distinguishing feature are spatially separated from each other (i.e., tail vs. mouth), and attention is most likely allocated to the relevant region within the fish shape. In Experiment 1B, in contrast, the distinguishing and non-distinguishing features are spread over the entire shape and are spatially overlapping (i.e., texture and colour). Possibly, in this experiment it is more difficult to ignore the irrelevant feature dimension, since participants engage in a spread attention mode, and fail to filter out spatially overlapping (yet irrelevant) information.

Taken together, the results of Experiment 1 show that priming by the fish object facilitated “same” responses to test pairs similar to its distinguishing feature relative to test pairs similar to its non-distinguishing feature. These results suggest that the distinguishing feature was primed, both when the distinguishing features were local features (Experiment 1a) and when they were surface features (Experiment 1b), thus supporting our hypothesis that during the process of object recognition attention is allocated to distinguishing features. Alternatively, it can be argued that the category (or type) of the fish object was primed, in which case the facilitated responses suggest that the internal representation of the object category (or type) involves some representation of distinguishing features. By this alternative interpretation, distinguishing features play a role in object recognition, but not necessarily via the allocation of attention to these features. Experiment 2 was designed to provide converging evidence for the hypothesis that attention is allocated to distinguishing features in the course of object recognition.

## Experiment 2

This experiment examined the hypothesis that attention is allocated to distinguishing features in the course of object recognition, by measuring attentional effects using a variant of the visual probe method (Navon & Margalit, 1983). This experiment also informs us whether attention to the distinguishing feature involves a spatial component.

The primary task was fish categorization (as in Experiment 1a). On some of the trials an intermediate-contrast probe was presented simultaneously with the fish, equally often near the distinguishing and the non-distinguishing features. If in the course of object recognition attention is allocated to the location of the distinguishing features, then probe detection rate should be higher when the probe appears near a distinguishing feature than near a non-distinguishing feature.

### *Method*

*Participants.* Nine students at the University of Haifa participated in the experiment. All had normal or corrected-to-normal vision.

*Stimuli.* The set of fish to be categorized was the same as in Experiment 1a (Figure 2a). In one condition the distinguishing feature was the shape of the tail (tail condition), and in the other condition the distinguishing feature was the shape of the mouth (mouth condition). The probe was a small intermediate contrast (0.5) grey (RGB 128, 128, 128) disk, subtending  $0.5^\circ$  in diameter.

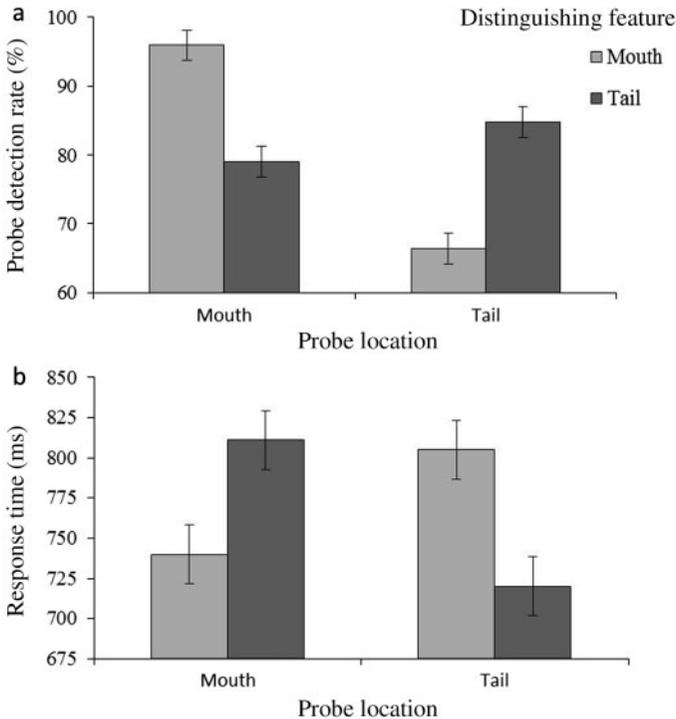
*Design and procedure.* The primary task in this experiment was the categorization of a fish, similar to the one in Experiment 1a. On 20% of all trials a probe was presented simultaneously with the fish, adjacent to the fish outline, equally often near the distinguishing and the non-distinguishing feature. The fish (and probe when present) were displayed for 200 ms. Categorization responses were made by pressing one of two keys (different fingers of the same hand); if a probe was detected, participants were required to double press the appropriate categorization key, making a combined response for the object categorization and the probe detection. This double press response method reduces any potential conflict that might arise when providing two separate responses at once: there cannot be any confusion regarding the order of responses, or as to which keys should be pressed—the selected categorization key is either pressed once indicating the categorization response alone, or double pressed (similar to a mouse double-click), indicating both the categorization decision and that the probe was detected. The participants were instructed to respond as quickly and as accurately as possible.

The experiment employed a 2 (distinguishing feature)  $\times$  2 (probe location) within subject design. The two distinguishing feature conditions—the mouth condition and the tail condition—were administered in separate blocks of 400 trials each, preceded by 30 practice trials. Each block began with a training phase for the categorization task, and the order of the blocks was counter balanced across participants.

### *Results and discussion*

*Categorization.* Performance in the categorization task was highly accurate (mean accuracy = 95.9%). Participants correctly recognized (categorized) the fish on 95.6% of the trials in the mouth condition and on 96.2% of the trials in the tail condition. Mean RT was 657 ms and 631 ms for the mouth and tail conditions, respectively.

*Probe detection.* Probe detection rate was calculated for all “probe present” trials on which the categorization response was correct. The mean percentage of probes detected was 79.6%, and the mean percentage of false alarms was 3.4%. Probe detection rates as a function of probe location and distinguishing feature are presented in Figure 6a. The 2 (distinguishing feature: mouth, tail)  $\times$  2 (probe



**Figure 6.** Results from Experiment 2. (a) Probe detection rate and (b) mean RT on probe detection trials, as a function of probe location and distinguishing feature.

location: mouth, tail) repeated measures ANOVA, performed on the arcsine transformed probe detection rates, showed a significant effect of probe location,  $F(1,8) = 16.87$ ,  $MSE = 12.24$ ,  $p < .01$ ,  $\eta_p^2 = .68$ . The percentage of probes detected was higher (by 11.8%) when probes appeared next to the mouth than next to the tail. The effect of distinguishing feature was non-significant,  $F < 1$ . Importantly, there was a significant interaction between distinguishing feature and probe location,  $F(1,8) = 20.66$ ,  $MSE = 23.99$ ,  $p < .01$ ,  $\eta_p^2 = .72$ . Detection rate of probes that appeared next to the mouth was higher (by 16.9%) when the distinguishing feature was the mouth than when it was the tail,  $F(1,8) = 6.8$ ,  $MSE = 52.47$ ,  $p < .05$ ,  $\eta_p^2 = .46$ . Similarly, when the probe appeared next to the tail, detection rate was higher (by 18.4%) when the distinguishing feature was the tail than when it was the mouth,  $F(1,8) = 8.31$ ,  $MSE = 39.64$ ,  $p < .05$ ,  $\eta_p^2 = .51$ .

We also examined the RT on the probe detection trials. On these trials the RT measure reflects the time needed both to categorize the fish and to detect

the probe.<sup>1</sup> Assuming that the difficulty of fish categorization is equivalent for all probe locations, any differences in the RT between the different probe locations can be attributed to the effect of probe location (near the distinguishing or the non-distinguishing feature) on the difficulty of probe detection.

RTs as a function of probe location and distinguishing feature are presented in Figure 6b. The 2 (distinguishing feature: mouth, tail)  $\times$  2 (probe location: mouth, tail) repeated measures ANOVA conducted on the RTs supported the probe detection rate results. There were no significant effects of distinguishing feature and of probe location ( $F_s < 1$ ), but the interaction between distinguishing feature and probe location was significant,  $F(1,8) = 19.54$ ,  $MSE = 2809.0$ ,  $p < .01$ ,  $\eta_p^2 = .72$ . When the probe appeared next to the mouth it was detected faster (by 65 ms) when the distinguishing feature was the mouth than when it was the tail. Similarly, when the probe appeared next to the tail it was detected faster (by 91 ms) when the distinguishing feature was the tail than when it was the mouth,  $F(1,8) = 4.13$ ,  $MSE = 9071.0$ ,  $p = .077$ ,  $\eta_p^2 = .34$ .

Thus, the results for the probe detection rate and the results for the RT on probe detection trials indicate that detection rate was higher and detection was faster for probes that appeared near the distinguishing than the non-distinguishing feature, suggesting that attention was allocated to the distinguishing features during the process of fish recognition. One might be concerned that the probe attracted attention in a bottom-up manner but, even if so, this would not explain the higher detection rate of the probe near the location of the distinguishing feature. The results of this experiment further suggest that the deployment of attention to distinguishing features involves a spatial component.

### Experiment 3

The purpose of Experiment 3 was to provide evidence that directing attention to the distinguishing features is not a result of recognizing an object, but rather, a mechanism that serves the recognition process. To this end, this experiment examined the effect of advance allocation of attention to the location of the distinguishing or the non-distinguishing feature on object recognition (categorization), using the spatial cueing method (e.g., Posner, Nissen, & Ogden, 1978). On each trial a transient pre-cue was presented at the location of either the distinguishing or the non-distinguishing feature, and then a fish was displayed for categorization. If indeed attention to distinguishing features facilitates object recognition, then categorization should be faster and/or more accurate when

---

<sup>1</sup> The combined recognition and probe-detection response was indicated by a fast double-press of the relevant recognition key. For the purpose of the RT analyses on probe-detection trials, we treated the interval between the two key presses as noise, using RT to the initial key press as the dependent measure.

attention is allocated in advance to the location of the distinguishing feature than to the location of the non-distinguishing feature (see Tsal & Kolbet, 1985).

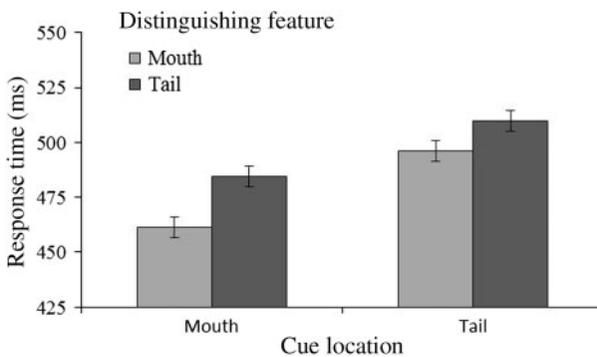
### Method

*Participants.* Nine students at the University of Haifa participated in the experiment. All had normal or corrected-to-normal vision.

*Stimuli.* The set of fish for the categorization task was similar to the one in Experiments 1a and 2 (Figure 2a), but the fish were longer (aspect ratio was 1.5:1 with respect to the fish in the previous experiments), increasing the distance between the locations of the mouth and tail. This change was made in order to ensure enough physical separation between the two pre-cued locations. As in Experiments 1a and 2, the distinguishing feature in one condition was the shape of the tail and in the other condition the shape of the mouth. The transient pre-cue was a small high-contrast white disk (RGB 255, 255, 255) with a diameter of  $0.5^\circ$ .

*Design and procedure.* Each trial began with the presentation of a small fixation cross for 500 ms followed by a transient pre-cue that was presented at one of two predefined locations: either next to the location of the distinguishing feature or next to the location of the non-distinguishing feature. The transient pre-cue was presented for 30 ms. Following a 100-ms ISI a fish was presented for 70 ms for categorization. The fish was facing left or right equally often. The participants had to identify the fish as quickly and as accurately as possible.

The experiment employed a 2 (distinguishing feature)  $\times$  2 (cue location) within participants design. The distinguishing feature conditions (mouth, tail) were administered in separate blocks and their order was counter balanced across participants. As in the previous experiments there was a training phase before



**Figure 7.** Results from Experiment 3. Mean RT in the fish categorization task as a function of cue location for the mouth and tail conditions.

each categorization task. Each block consisted of 400 trials preceded by 30 practice trials. Cue location conditions occurred on an equal number of trials and their order was randomized.

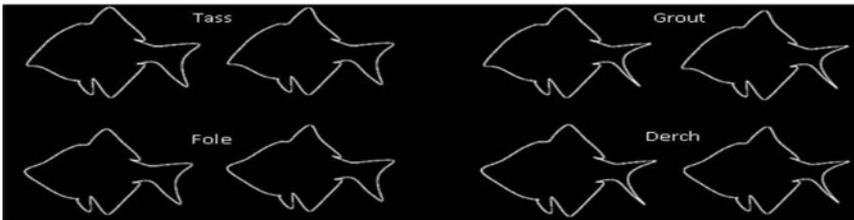
### Results and discussion

Performance in the categorization task was highly accurate (mean accuracy = 96.4%). Participants correctly recognized (categorized) the fish on 96.7% of the trials in which the distinguishing feature was cued and on 96.1% of the trials in which the non-distinguishing feature was cued.

All RT summaries and analyses are based on participants' mean RTs for correct categorization responses. RTs shorter than 200 ms and longer than 2000 ms were discarded (0.2% of all trials).

Mean RTs for the cue location conditions as a function of distinguishing feature are presented in Figure 7. A 2 (distinguishing feature: mouth, tail)  $\times$  2 (cue location: distinguishing feature, non-distinguishing feature) ANOVA performed on the RT data showed that participants were faster in recognizing the fish when the distinguishing feature was the mouth than when it was the tail, as indicated by the significant effect of distinguishing feature,  $F(1,8) = 8.58$ ,  $MSE = 906.0$ ,  $p < .05$ ,  $\eta_p^2 = .52$ . Importantly, there was a significant effect of cue location,  $F(1,8) = 27.08$ ,  $MSE = 119.1$ ,  $p = .001$ ,  $\eta_p^2 = .77$ , but the interaction between distinguishing feature and cue location was not significant,  $F(1,8) = 1.02$ ,  $MSE = 163.2$ ,  $p = .34$ ,  $\eta_p^2 = .11$ . As can be seen in Figure 7, RT was faster when the location of the distinguishing feature was cued than when the location of the non-distinguishing feature was cued, both in the mouth condition (by 24 ms),  $F(1,8) = 52.26$ ,  $MSE = 46.48$ ,  $p < .001$ ,  $\eta_p^2 = .87$ , and in the tail condition (by 15 ms),  $F(1,8) = 4.09$ ,  $MSE = 235.85$ ,  $p = .078$ ,  $\eta_p^2 = .34$ .

These results indicate that allocating spatial attention in advance to the location of the distinguishing features can facilitate object recognition performance. Taken together with the results of the preceding experiments, indicating that attention is in fact differentially allocated to distinguishing features during



**Figure 8.** Examples of the fish stimuli used in Experiment 4. A given fish is distinguished from the other fish by different features, depending on the stimulus set. For example “Tass” differs from “Grout” by its tail but from “Fole” by its mouth.

object recognition, the simplest conclusion is that this allocation serves the recognition process.

## DISTINGUISHING FEATURES ARE CONTEXT-DEPENDENT

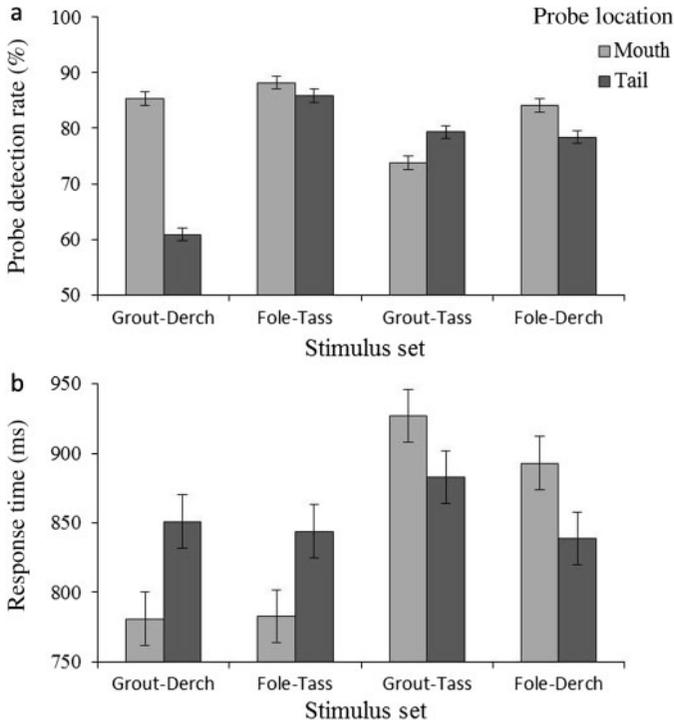
### Experiment 4

The results of the preceding experiments provide converging evidence for the basic claim that, in the course of object recognition, visual attention (both spatial and feature-based) is allocated to distinguishing features. Experiment 4 was designed to test an auxiliary hypothesis regarding the nature of the distinguishing features themselves. By definition, a distinguishing feature is one that is diagnostic of object identity. However, a particular feature may be diagnostic in one context but not in another. For example, a feature that is diagnostic for identifying a wolf in a herd of sheep may be of no use at all for identifying a wolf in a herd of wild dogs. Hence, the distinguishing features, to which attention is allocated in the course of object recognition, are expected to be *context dependent*: They should be used to recognize a particular object only in contexts in which the feature is diagnostic of object identity.

Once it has been established that attention is directed to distinguishing features (Experiments 1–3), attentional effects, measured by the visual probe method, can be used to examine the context-dependence hypothesis, by indicating which features are being utilized as distinguishing features in different contexts.

Participants learned four fish categories, each defined by a combination of two features: the shape of the mouth and the shape of the tail (see [Figure 8](#)). Context was manipulated by stimulus set. Each of the four stimulus set conditions included two categories such that each fish category participated in two different stimulus sets. For example, in one condition, Grout had to be identified in the context of Tass (i.e., Grout-Tass stimulus set), in which the diagnostic feature is the tail, and in another condition, Grout had to be identified in the context of Derch (i.e., Grout-Derch stimulus set), in which the diagnostic feature is the mouth. Manipulating the stimulus set in this way models ecological situations in which an object is recognized in different environments.

The probe detection paradigm was used, and probe detection rate at different feature locations was taken to indicate which of the features were being attended to and, hence, being used as distinguishing features: higher detection rate at one feature location than at another feature location indicates that the former is being used as the distinguishing feature. If indeed distinguishing features are context dependent, then an interaction between stimulus set and probe location is expected, such that for a given fish category, detection rate for probes at a given feature location (mouth or tail) varies with the stimulus set. For example, the highest probe detection rate is expected to be at a different location in the



**Figure 9.** Results from Experiment 4. (a) Probe detection rate and (b) RT on probe detection trials, as a function of probe locations, for each stimulus set. See text for details.

Grout-Tass condition (next to the tail) than in the Grout-Derch condition (next to the mouth), because tail is diagnostic of object identity in the former condition, whereas mouth is diagnostic in the latter condition.

### Method

*Participants.* Eleven students at the University of Haifa participated in the experiment. All had normal or corrected-to-normal vision.

*Stimuli.* The set of fish in this experiment belonged to four categories, differing from each other by a specific diagnostic feature—the shape of the mouth or the shape of the tail. The dorsal fin of the fish in all categories was either straight or concave and thus non-diagnostic (Figure 8). The probe was a grey disk (RGB 128, 128, 128) of intermediate contrast (0.5), and subtended  $0.5^\circ$  in diameter.

*Design and procedure.* The experiment began with the presentation of the four fish categories (Figure 8) followed by a general learning phase. Each learning trial began with the presentation of a small fixation cross at the centre of the

screen for 500 ms and then a white outline of a fish was displayed on a black background until response. The participants had to decide as quickly and accurately as possible to which of the four categories it belonged by pressing one of four keys. Participants were required to press the “1” and “2” keys on the right hand side of the keypad to indicate the Tass and Grout categories, respectively, using two right hand fingers, and the “s” and “d” keys on the left side of the keypad to indicate the Derch and Fole categories, respectively, using two left hand fingers. A short audio feedback was provided when the participant pressed a wrong key. The general learning phase lasted until the participant had at most two errors in 20 consecutive trials.

The general learning phase was followed by four blocked experimental conditions, each defined by the specific stimulus set: Grout-Tass (diagnostic feature T), “Grout-Derch” (diagnostic feature M), “Fole-Derch” (diagnostic feature T), and Fole-Tass (diagnostic feature M). The order of the four blocked conditions was counterbalanced across participants. Each condition started by presenting the relevant stimulus set (i.e., the two fish categories included in that condition) followed by a training phase, which was similar to the general learning phase except for the number of response options, and that now on 40% of the trials a probe appeared equally often at one of two locations: near the mouth or near the tail. When a probe was detected participants were required to perform the categorization task by a double press on the appropriate key. The training phase was followed by the experimental condition, in which the fish was presented for 200 ms and a probe appeared on 20% of the trials. Each experimental condition included 400 trials, preceded by 30 practice trials.

### *Results and discussion*

*Categorization task.* Performance in the categorization task was highly accurate (mean accuracy = 95.9%). Participants correctly recognized (categorized) the fish on 95.3%, 96.8%, 95.8% and 96.85% of the trials in the Grout-Tass, Tass-Fole, Fole-Derch and Grout-Derch conditions, respectively. Mean RT in these conditions was 901 ms, 845 ms, 895 ms and 865 ms, respectively.

*Probe detection.* Probe detection rates were calculated for all trials on which there were no recognition (categorization) errors. Mean probe detection rate was 77.5%, and the mean percentage of false alarms was 1.7%. Mean probe detection rates as a function of probe location and stimulus set are presented in [Figure 9a](#). The 4 (stimulus set)  $\times$  2 (probe location) repeated measures ANOVA, performed on the arcsine transformed probe detection rates, showed a significant effect of stimulus set,  $F(3,30) = 4.3$ ,  $MSE = 36.84$ ,  $p < .01$ ,  $\eta_p^2 = .3$ , with the highest average detection rate in the Fole-Tass condition and the lowest in the Grout-Derch condition. Most importantly, the analysis revealed, as predicted, a significant interaction between stimulus set and probe location,  $F(3,30) = 7.65$ ,

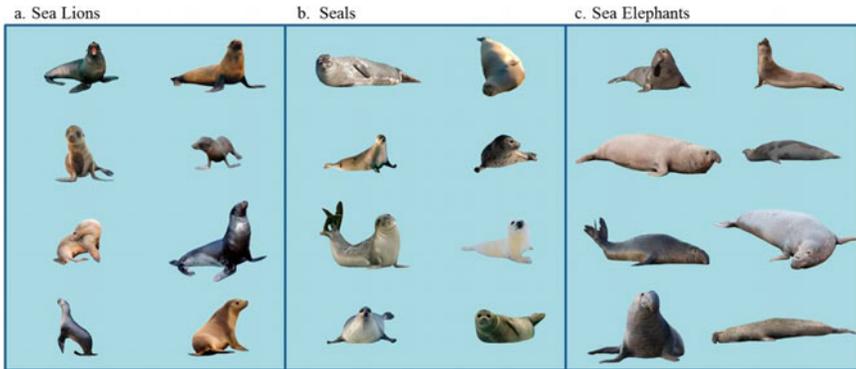


Figure 10. Examples of the stimuli that were used in Experiment 5. (a) Sea lions (b) Seals (c) Sea elephants.

MSE = 16.97,  $p < .001$ ,  $\eta_p^2 = .43$ , indicating that probe detection rate at the different locations depended on the stimulus set. Thus, when identifying Grout vs. Derch, probe detection rate was higher for probes that appeared next to the mouth than next to the tail (by 24.4%,  $F(1,10) = 14.51$ , MSE = 58.5,  $p < .01$ ,  $\eta_p^2 = 0.59$ ), whereas the opposite occurred when identifying Grout vs. Tass probe detection rate was higher (by 5.4%) for probes that appeared next to the tail than next to the mouth and, in turn, when identifying Tass vs. Fole probe detection rate was higher (by 2.4%) for probes that appeared next to the mouth than next to the tail. We expected the opposite to occur when identifying Fole vs. Derch, but the results did not match the expectations: probe detection rate was higher (by 5.7%) when the probe was next to the mouth than when it was next to the tail. Notwithstanding this latter exception, the results showed that for a given fish category, the location for which the detection rate of the probe was the highest depended on the stimulus set.

Similar results were observed for the RTs on the probes detection trials. RTs as a function of probe location and stimulus set are presented in Figure 9b. The 4 (stimulus set)  $\times$  2 (probe location) repeated measures ANOVA showed a significant effect of stimulus set,  $F(3,30) = 5.18$ , MSE = 8181.2,  $p < .01$ ,  $\eta_p^2 = .34$ , with the fastest performance in the Fole-Tass condition and the slowest in the Grout-Tass condition. Most importantly, the analysis revealed a significant interaction between stimulus set and probe location,  $F(3,30) = 6.65$ , MSE = 3694.4,  $p = .001$ ,  $\eta_p^2 = .40$ , indicating that probe detection time at the different locations depended on the stimulus set. When identifying Grout vs. Derch and Fole vs. Tass, detection was faster (by 70 ms,  $F(1,10) = 5.43$ , MSE = 5020.2,  $p < .05$ ,  $\eta_p^2 = .35$ , and 61 ms,  $F(1,10) = 6.19$ , MSE = 3378.4,  $p < .05$ ,  $\eta_p^2 = .38$ , respectively) when the probe appeared next to the mouth than next to the tail, whereas when identifying Grout vs. Tass and Fole vs. Derch, detection was faster (by 44 ms,  $F(1,10) = 7.83$ , MSE = 1395.1,  $p < .05$ ,  $\eta_p^2 = .44$ , and 54 ms,

$F(1,10) = 3.95$ ,  $MSE = 4074.3$ ,  $p = .075$ ,  $\eta_p^2 = .28$ , respectively) when the probe appeared next to the tail than next to the mouth.

These RT results converge with the accuracy data in indicating that for any given fish category (object type), the distinguishing feature used to identify a particular fish as an instance of that category (object type) was not fixed, but rather varied with the stimulus set. These results support the idea that distinguishing features are context dependent.

## NATURAL OBJECTS

### Experiment 5

Experiments 1–4 used two-dimensional line drawings to examine allocation of attention when a single, well-defined feature can be used to recognize the objects. Although it is common practice to draw conclusions regarding object categorization and object recognition in tasks using line drawings as the object stimuli (e.g., Biederman, 1987; Humphrey & Jolicoeur, 1993; Sigala & Logothetis, 2002), the exclusive use of object line drawings raises the issue of whether the results generalize to more natural objects, which are inherently more “noisy”—both in the overall visual data and in the distinguishing features themselves. To address this issue, the object stimuli used in this experiment were photographs (taken mostly from the Internet) of real sea mammals, specifically, seals, sea lions and sea elephants. The pictures were cropped so that each animal appeared on a homogeneous standard background. None of the animals were physically identical; they were photographed in a variety of many different postures, from different distances and viewpoints, and under different lighting conditions. All of these factors contributed to large variability in visual information, both between and within categories (see Figure 10 for examples).

The experiment involved two stimulus sets, presented in two different blocks, each of which included only two of the four types of sea animals: one set included seals and sea elephants and the other set included seals and sea lions. Thus, seals participated in two different stimulus sets. In one block, the participants had to decide whether a presented animal is a seal or a sea elephant and in the other block, whether it is a seal or a sea lion.

Given the context dependency of distinguishing features (based on the results of Experiment 4), we expected that different distinguishing features will be utilized in the different sets for discriminating the seals from the other animals in the set. Seals differ from sea elephants by the shape of their nose. They differ from sea elephants also in their size; however it is difficult to tell the size of an object from a photograph when there is no background or other reference to compare to. Seals and sea lions differ formally by two features: ears (sea lions have small outer ears whereas the ears of the seals are invisible) and hind legs (sea lions have hind legs with a joint, so they can bend their legs forward and utilize them for

movement whereas hind legs of seals are degenerated), but the hind legs are bigger, easier to see and more salient than the small and barely seen ears of the sea lions. Thus, we hypothesized that the diagnostic feature for discriminating the seal from the sea elephant would be the nose, whereas the diagnostic feature for discriminating the seal from the sea lion would be the hind legs. Therefore, it was expected that when the animal presented for recognition (categorization) is a seal, attention, measured by a probe detection task, would be directed to the nose in the seal–sea elephant condition, and to the hind legs in the seal–sea lion condition. Note, that participants did not receive any explicit instructions or explanations regarding the features that distinguish seals from sea lions or from sea elephants.

### *Method*

*Participants.* Eight students at the University of Haifa participated in the experiment. All had normal or corrected-to-normal vision.

*Stimuli.* The set of objects to be recognized in this experiment were instances of three types of sea animals: seals, sea lions and sea elephants (Figure 10). The most salient diagnostic feature by which seals differ from sea elephants is their nose. The most salient feature by which seals differ from sea lions is their hind legs. There are additional features that distinguish these animals from each other (Figure 10). The animals were placed on a homogeneous bright blue background (RGB 182, 216, 225), subtending  $15^\circ \times 9^\circ$ . There were 40 images for each category. A low contrast (0.08), small round probe (RGB 167, 199, 207), subtending  $0.5^\circ$  in diameter was used for the probe detection task.

*Design and procedure.* There were two experimental conditions defined by the stimulus set: “seals–sea lions” and “seals–sea elephants”. The two conditions were administered in separate blocks, the order of which was counterbalanced across participants. Each block began with the presentation of instances of the two types of sea animals to be recognized in that block, followed by a training phase. Twenty instances of each type of animal were used in the training phase. Each training trial began with the presentation of a small fixation cross at the centre of the screen for 500 ms. It was followed by the display of a single animal on a blue rectangular background, centred on the black screen. The animal was displayed for 1000 ms. The blue rectangle was left on the screen for additional 1000 ms. In 66% of the trials, a small low contrast probe was displayed on the blue rectangle. The probe could be seen as soon as the animal disappeared, as if it was “behind” it, and it disappeared together with the blue rectangular background, to prevent capture of attention by abrupt onset or offset. The probe appeared equally often at the location of the previously displayed nose of the animal or at the location of the previously displayed hind legs of the animal. Note that because the animal could be photographed at any posture, the actual location of the probe on the blue rectangle could vary greatly. The participants

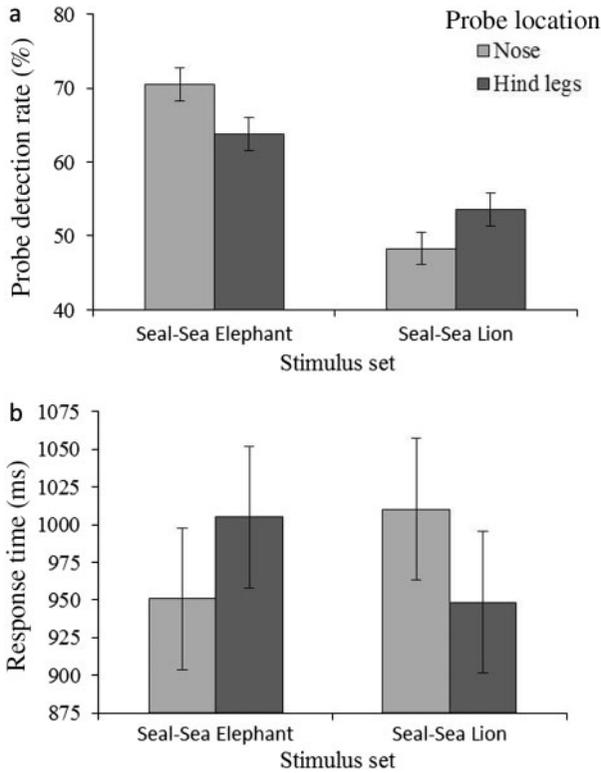


Figure 11. Results from Experiment 5. (a) Probe detection rate and (b) Mean RT, as a function of probe location and stimulus set when identifying a seal.

had to decide as quickly and accurately as possible which type of animal was presented by pressing one of two keys (different fingers of the same hand); if the probe was detected, participants were required to double press the appropriate categorization key, making a combined response for the object categorization and the probe detection. A short audio feedback was provided when the participant pressed a wrong key. The training phase lasted until the participant had at most two errors in 20 consecutive trials.

In the experimental phase that followed the training phase the animal was presented for 200 ms, the blue rectangle was displayed for an additional duration of 200 ms and the probe appeared on 33% of the trials. Thirty instances of each animal category were used in the experimental phase, 10 of which were also used in the training phase. Each experimental condition included 360 trials, preceded by 30 practice trials.

### Results and discussion

*Categorization task.* Performance in the categorization task was highly accurate (mean accuracy = 93%, in both conditions). Mean RT in the seals–sea lions and seals–sea elephants conditions was 965.7 ms and 868.9 ms, respectively. This numerical difference, however, was not statistically significant,  $F(1,7) = 1.3$ ,  $MSE = 29519.3$ ,  $p = .29$ ,  $\eta_p^2 = .15$ .

*Probe detection.* Probe detection rates were calculated for all trials on which the sea animal was identified correctly. The mean percentage of probes detected was 56%, and the mean percentage of false alarms was 6.3%. We examined probe detection rates when identifying a seal as a function of probe location and stimulus set (see Figure 11). The 2 (stimulus set: seals–sea lions, seals–sea elephants)  $\times$  2 (probe location: nose, hind legs) repeated measures ANOVA, performed on the arcsine transformed probe detection rates showed, as expected, a significant interaction between stimulus set and probe location,  $F(1,7) = 8.18$ ,  $MSE = 2.4$ ,  $p < .05$ ,  $\eta_p^2 = .54$ . The ANOVA conducted on the RT data did not yield a statistically significant interaction,  $F(1,7) = 1.03$ ,  $MSE = 26058.1$ ,  $p = .34$ ,  $\eta_p^2 = .13$ . When the stimulus set was “seals–sea elephants”, probe detection rate was higher (by 7.7%;  $t(7) = 1.46$ ,  $p = .088$ ,  $\eta_p^2 = .23$ , one-tailed) and RT was faster (by 54.3 ms) at the nose than at the hind legs. Similarly, when the stimulus set was “seals–sea lions”, detection rate was higher (by 5.3%;  $t(7) = 1.70$ ,  $p = .06$ ,  $\eta_p^2 = .29$ , one-tailed) and RT was faster (by 61 ms) at the hind legs than at the nose. These results indicate that, as expected, attention was allocated to the distinguishing feature that was relevant for recognizing the animal in a given context: it was allocated either to the nose or to the hind legs of the seal, depending on the stimulus set.

## GENERAL DISCUSSION

The present work was designed to provide systematic evidence for a particular role of visual attention in object recognition. Specifically, we examined the hypothesis that in the course of object recognition, attention is directed to distinguishing features—those that are most diagnostic of object identity in a particular context.

The first three experiments (Experiments 1–3) provided converging evidence for the claim that attention is allocated to the distinguishing features in the course of object recognition, using two-dimensional line drawings in an object categorization task. Experiment 1 showed that categorizing a fish primed its distinguishing features but not its other features—both local contour features (mouth, tail; Experiment 1a) and surface features (texture, colour; Experiment 1b). Experiment 2 showed that when categorizing a fish, probe detection rate was higher and response time was faster for probes appearing near the distinguishing feature than for probes appearing near the non-distinguishing

feature. Finally, Experiment 3 showed that advance allocation of attention to the location of the distinguishing feature by a transient cue yielded faster fish categorization than when the location of a non-distinguishing feature was pre-cued.

Experiments 4 and 5 showed that the attended distinguishing features are context dependent: Attention was allocated to different features of the same object (“Derch” and “Tass” fish in Experiment 4; seals in Experiment 5), depending on the overall stimulus set. Moreover, using photographs of natural objects as stimuli, Experiment 5 showed that the allocation of attention to distinguishing features in a context-dependent manner also holds for the categorization of pictures of natural objects (sea animals) despite substantial within-category variation in physical characteristics, posture, viewing angle, and more.

## Object recognition and categorization

The focus of the current study is on the role of attention in object recognition. We investigated this issue at the operational level using object categorization tasks, which allow greater control over the object features, both distinguishing and non-distinguishing. In the view that guided our work, object recognition is conceived as a process of discrimination between probable alternatives (Gregory, 1966; von Helmholtz, 1867), an iterative process in which bottom-up and top-down processes interact, and in which attention plays a critical role. By this view, object recognition is essentially a matter of categorization, as the set of possible alternatives for any particular visual stimulus is always limited by one’s expectations in a specific context. Thus, both object recognition and object categorization tasks (and all computational models of object recognition; see next section) involve identifying a presently viewed visual stimulus as a member of a particular object category (see also Meuwese, van Loon, Lamme, & Fahrenfort, 2014).

Typically, everyday situations evoke expectations about probable objects, based on world knowledge, context, and goals (e.g., Norman & Bobrow, 1976; Palmer, 1975). For example, when looking up and wondering what is that object in the sky, our expectations dictate a very limited set of alternatives. Similarly, in the laboratory, expectations are evoked by the experimental task. Pure data-driven recognition—where an object could be anything—is presumably quite rare, and can be seen as a special case in which the probable alternatives are all objects known to the observer. Therefore, we believe that the findings in the tasks used in our experiments can be interpreted directly in terms of the role of attention to distinguishing features in the process of object recognition.

Nonetheless, we acknowledge that the specific processes and strategies used to recognize objects may depend, in part, on the size of the effective stimulus set, and also perhaps in laboratory settings be influenced by the fact that the

recognition/categorization decisions are made repeatedly for a large number of trials. These are issues that pertain to all experimental research on object recognition, and we point to the need to follow up on the present findings using larger stimulus sets that include a larger number of possible alternative objects for recognition (e.g., Baruch, Kimchi, & Goldsmith, *in press*).

### Attention to distinguishing features in object recognition

In this work, we advanced the idea that attention plays a special role in object recognition, namely, it serves to facilitate the extraction of diagnostic object information. As noted earlier, the information that can be extracted from the visual scene at a glance, in a completely data-driven manner, may often be inadequate for object recognition. In such cases, the present study suggests that attention to distinguishing features is required. Expectations regarding the identity of the object (e.g., it is either a seal or a sea elephant) drive the allocation of attention to distinguishing features—any type of visual information that is diagnostic of object identity in a specific context—leading to the extraction of additional or higher quality information, providing further constraints that reduce uncertainty and refine the hypotheses (e.g., “it is a seal”). Additional results from our laboratory reported elsewhere (Baruch et al., *in press*; see also Blair et al., 2009) show that this process is repeated, if necessary, in an iterative manner until recognition is achieved.

Attention to distinguishing features may have computational advantages as well. Feature-based models of object recognition generally represent objects as points in some  $n$ -dimensional feature space (e.g., Duda & Hart, 1973; Pontil & Verri, 1998). An object category may be represented by some measure of the distribution of its instances (e.g., central tendency, variability), or by a “typical example” or “prototype” (e.g., Rosch & Mervis, 1975). By this approach, an object is recognized by assigning it to the nearest category, as calculated by some distance measure. The main problem faced by such models derives from the inherent conflict between sensitivity and stability (e.g., Edelman, 1997; Marr, 1982), the ability to reliably distinguish between visually similar yet different objects belonging to different categories, and at the same time recognize the same type of object despite within-category variation in visual properties, which is exacerbated by differences in viewpoint, illumination, and so forth. Attention to distinguishing features may be viewed as a mechanism for reducing the sensitivity-stability trade-off by: (a) reducing the dimensionality of the feature space by focusing on fewer features, (b) maximizing the distance between competing representations, and (c) improving the quality (e.g., resolution; Carrasco & Yeshurun, 1998) of the visual-featural information.

As implied in such models, in many real-life situations there are larger numbers of possible objects and object categories than used in our experiments, and it may often be necessary to base the recognition decision on more than a

single distinguishing feature. Conceivably, attention could be allocated in parallel or in a serial manner. However, there may be cases in which the deployment of attention must be sequential in order to achieve recognition, for example when the distinguishing features are subtle and spatially disparate. Reiterative sequential allocation may also be needed as hypotheses regarding object identity are refined and tested (e.g., Baruch et al., *in press*; Blair et al., 2009).

Another interesting case is when there are multiple distinguishing features, any one of which is sufficient for recognition, but differing in their perceptual discriminability. In such a case, one might expect a tendency to direct attention to the feature that is most easily discriminable. Consider, for example, the problem of recognizing an animal as a seal or a sea lion in Experiment 5. Although these sea animals can be distinguished either by their ears (external versus internal) or by their hind legs (with versus without a joint), attention was in fact allocated to the hind legs—the distinguishing feature that was easier to discriminate.

The distinguishing features that one attends to, and the ease of attending to them, might also depend on experience. This is implied by research on the object categorization performance of expert versus novice observers. Whereas novice observers perform basic-level categorization tasks (bird vs. fish) faster than subordinate level categorizations (robin vs. sparrow; trout vs. bass), there is no difference in the time needed by experts to perform these categorization tasks (e.g., Gillebert, Op de Beeck, Panis, & Wagemans, 2008; Tanaka & Taylor, 1991). In explaining this pattern of results, some researchers have suggested that in contrast to the novice observers, the visual systems of expert observers have adapted such that they can detect the fine-detailed features needed to perform the subordinate categorization tasks without focused attention (e.g., Gillebert et al., 2008; Van Rullen, Reddy, & Koch, 2004). However, it has also been suggested that expert observers know what to look for—that category learning involves learning to attend optimally (e.g., Lamberts, 1995; Nosofsky, 1984).

### Concluding remarks

The results of this study reinforce and extend previous findings, suggesting that attention is directed to distinguishing features in the course of object recognition—at least when the diagnostic information cannot be obtained at a glance. Future research will need to investigate the specific conditions under which attention is allocated to distinguishing features, and the underlying mechanisms. Any comprehensive model of the object recognition process will need to account for these findings.

## REFERENCES

- Aha, D. W., & Goldstone, R. L. (1990). *Learning attribute relevance in context in instance-based learning algorithms*. Proceedings of the Twelfth Annual Conference of the Cognitive Science Society. (pp. 141–148). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*, 600–609. doi:10.1126/science.8316836
- Baruch, O., Kimchi, R., & Goldsmith, M. (in press). *Object recognition is an interactive iterative process*.
- Beller, H. K. (1971). Priming: Effects of advance information on matching. *Journal of Experimental Psychology, 87*, 176–182. doi:10.1037/h0030553
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147. doi:10.1037/0033-295X.94.2.115
- Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 1196–1206. doi:10.1037/a0016272
- Broadbent, D. E. (1958). *Perception and communication*. New York, NY: Oxford University Press.
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York, NY: Wiley.
- Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences, 1*, 296–304. doi:10.1016/S1364-6613(97)01090-5
- Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1476–1492. doi:10.1037/0096-1523.31.6.1476
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Perception Science, 2*, 243. doi:10.3389
- Ganis, G., Schendan, H. E., & Kosslyn, S. M. (2007). Neuroimaging evidence for object model verification theory: Role of prefrontal control in visual object categorization. *Neuroimage, 34*, 384–398. doi:10.1016/j.neuroimage.2006.09.008
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton-Century-Crofts.
- Gillebert, C. R., Op de Beeck, H. P., Panis, S., & Wagemans, J. (2008). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. *Journal of Cognitive Neuroscience, 21*, 1054–1064.
- Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178–200. doi:10.1037/0096-3445.123.2.178
- Gregory, R. L. (1966). *Eye and brain*. New York, NY: World University Library.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791–804. doi:10.1016/S0896-6273(02)01091-7
- Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience, 8*, 197–230. doi:10.1037/0033-295X.8.2.3.184
- Humphrey, G. K., & Jolicoeur, P. (1993). An examination of the effects of axis foreshortening, monocular depth cues, and visual field on object identification. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 46*(1), 137–159. doi:10.1080/14640749308401070
- Johnson, J. S., & Olshausen, B. A. (2005). The recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Research, 45*, 3262–3276. doi:10.1016/j.visres.2005.06.007

- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, *111*, 880–913. doi:10.1037/0033-295X.111.4.880
- Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*, 161–180. doi:10.1037/0096-3445.124.2.161
- Li, F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Science, USA*, *99*, 9596–9601.
- Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*, 616–618. doi:10.1038/383616a0
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*, 657–672. doi:10.3758/BF03209251
- Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.
- Meuwese, J., Loon, A., Lamme, V., & Fahrenfort, J. (2014). The subjective experience of object recognition: Comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics*, *76*, 1057–1078.
- Navon, D., & Margalit, B. (1983). Allocation of attention according to informativeness in visual recognition. *The Quarterly Journal of Experimental Psychology Section A*, *35*, 497–512. doi:10.1080/14640748308402484
- Norman, D. A., & Bobrow, D. G. (1976). On the role of active memory processes in perception and cognition. In C. N. Cofer (Ed.), *The structure of human memory* (pp. 114–132). San Francisco, CA: W. H. Freeman.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57. doi:10.1037/0096-3445.115.1.39
- Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory–cognitive interaction. In D. A. Norman & D.E. Rumelhart (Eds.), *Explorations in cognition* (pp. 279–307). Hillsdale, NJ: Erlbaum.
- Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, *58*(1), 75–96. doi:10.1146/annurev.psych.58.102904.190114
- Pontil, M., & Verri, A. (1998). Support vector machines for 3D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 637–646. doi:10.1109/34.683777
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. doi:10.1080/00335558008248231
- Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role for spatial location. In N. H. L. Pick & I. J. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137–157). Hillsdale, NJ: Erlbaum.
- Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. doi:10.1016/j.cogpsych.2004.11.001
- Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 811–829.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233. doi:10.1037/0096-3445.104.3.192
- Rosch, E., & Mervis C. E. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605. doi:10.1016/0010-0285(75)90024-9
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*, 318–320. doi:10.1038/415318a
- Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457–482. doi:10.1016/0010-0285(91)90016-H

- Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 37, 571–590. doi:10.1080/14640748508400920
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. doi:10.1016/0010-0285(80)90005-5
- Tsal, Y., & Kolbet, L. (1985). Disambiguating ambiguous figures by selective attention. *The Quarterly Journal of Experimental Psychology*, 37(1), 25–37. doi:10.1080/14640748508400950
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352. doi:10.1037/0033-295X.84.4.327
- VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual-tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience*, 16(1), 4–14.
- von Helmholtz, H. (1867). *Treatise on physiological optics* (Volume 3, J. P. C. Southall, Trans.). New York, NY: Dover.
- Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision*, 8(4), 1–12.
- Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396, 72–75. doi:10.1038/23936