

Psychological Science

<http://pss.sagepub.com/>

When Categories Collide : Accumulation of Information About Multiple Categories in Rapid Scene Perception

Karla K. Evans, Todd S. Horowitz and Jeremy M. Wolfe
Psychological Science published online 9 May 2011
DOI: 10.1177/0956797611407930

The online version of this article can be found at:
<http://pss.sagepub.com/content/early/2011/05/09/0956797611407930>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Association for Psychological Science](http://www.sagepub.com)

Additional services and information for *Psychological Science* can be found at:

Email Alerts: <http://pss.sagepub.com/cgi/alerts>

Subscriptions: <http://pss.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

When Categories Collide: Accumulation of Information About Multiple Categories in Rapid Scene Perception

Karla K. Evans^{1,2}, Todd S. Horowitz^{1,2}, and Jeremy M. Wolfe^{1,2}

¹Brigham and Women's Hospital and ²Visual Attention Lab, Harvard Medical School

Psychological Science

XX(X) 1–8

© The Author(s) 2011

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797611407930

http://pss.sagepub.com



Abstract

Experiments have shown that people can rapidly determine if categories such as “animal” or “beach” are present in scenes that are presented for only a few milliseconds. Typically, observers in these experiments report on one prespecified category. For the first time, we show that observers can rapidly extract information about multiple categories. Moreover, we demonstrate task-dependent interactions between accumulating information about different categories in a scene. This interaction can be constructive or destructive, depending on whether the presence of one category can be taken as evidence for or against the presence of the other.

Keywords

nonselective processing, attention, scenes, gist, interactions

Received 9/10/10; Revision accepted 12/23/10

Humans can extract surprisingly complex semantic and statistical information from complex scenes that are presented very briefly or with limited focused attention. Within 200 ms, observers can assess the mean and distribution of size (Chong & Treisman, 2003), orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), and other basic visual attributes (Chubb, Nam, Bindman, & Sperling, 2007; Melcher & Kowler, 1999) for arrays of many objects without needing to attend to each object. Within 120 ms, observers are still able to identify aspects of the meaning, or “gist,” of a novel scene (e.g., picnic or birthday party; Potter & Faulconer, 1975), as well as to recognize small objects (Fei-Fei, Iyer, Koch, & Perona, 2007) or report their locations and spatial relations (Evans & Treisman, 2005; Tatler, Gilchrist, & Rusted, 2003). Even with shorter, masked viewing durations (19 to 50 ms), observers are able to classify a scene at basic (e.g., lake vs. forest) and superordinate (e.g., natural vs. urban) levels (Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007), and they can determine how pleasant it is (Kaplan, 1992). Furthermore, large objects can be identified (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001), even when focused attention is engaged with another foveal task (Li, VanRullen, Koch, & Perona, 2002). We refer to properties that can be reported accurately under such circumstances as *nonselective*, because they can apparently be perceived without directing selective attention to individual objects (even though selective attention to individual objects is required by similar, seemingly simpler tasks; Vickery, King, & Jiang, 2005).

The existence of such nonselective processing has inspired a set of feed-forward models of visual processing in which quite complex properties can be extracted from the initial sweep of neural activity from retina to cortex, without feedback or other top-down influences (Fukushima & Miyake, 1982; Itti & Koch, 2001; Riesenhuber & Poggio, 2000; Rousselet, Thorpe, & Fabre-Thorpe, 2004; Serre, Oliva, & Poggio, 2007; Thorpe, Delorme, & Van Rullen, 2001). Some researchers have extrapolated from such models and results to the conclusion that the visual system automatically computes multiple nonselective properties at the same time (Rousselet et al., 2004; Serre et al., 2007). This conclusion comports well with the introspective impression that the visual world is rich and detailed, even in a single glance.

However, this conclusion is not necessarily warranted by the data, and introspection is not definitive evidence. It is important to keep in mind that in most of the experiments just cited, observers were asked to report repeatedly on only one or two properties over many trials (e.g., Is there an animal in the scene?), and these properties were specified well in advance. One could explain the existing data equally well by assuming that only a small number of nonselective filters can be applied to the feed-forward visual data stream; although the number of

Corresponding Author:

Karla K. Evans, Harvard Medical School, Visual Attention Lab, 64 Sidney St., Suite 170, Cambridge, MA 02139

E-mail: kevens@search.bwh.harvard.edu

potential filters may be large, maybe they cannot all be active simultaneously. Indeed, one might imagine that trying to extract more than one nonselective property from the same brief image might lead to destructive interference between filters, such that selective attention to one property is required in order to disambiguate the information.

Can two or more properties be extracted in a single nonselective step? Suppose that two questions are asked about the same image: Is there an animal present? Is this a beach scene? It seems possible that the feed-forward sweep of information could provide answers for both of these questions. Alternatively, the system might need to be configured in advance to direct the information to one or the other classifier. Of course, one can also imagine intermediate states between these extremes. We conducted a series of experiments to investigate this issue. The results presented here support three conclusions: First, multiple nonselective calculations can occur simultaneously; second, these calculations interact with each other; and third, the nature of the interaction (i.e., whether it is constructive or destructive) depends on the structure of the task.

Experiment I

Can observers monitor a briefly presented scene for multiple properties simultaneously? Previous studies have employed only one or two predetermined target categories (e.g., animals or people). In Experiment 1, we employed nine different gist categories that could be used to label the stimulus scenes (animal, human, vehicle, bridge, flower, mountain, beach, street, or indoor scene). The cued category varied randomly from trial to trial, and cues were given in advance of the stimulus on some trials (precues) and after the stimulus on others (postcues).

Method

Observers. Ten observers (5 females, 5 males; age range = 19–42 years) were recruited from the Brigham and Women's Hospital subject pool. Each observer passed the Ishihara (1987) test for color blindness and had normal or corrected-to-normal vision. All observers gave informed consent, as approved by the Partners Healthcare Corporation Institutional Review Board and were compensated for their time.

Stimuli and apparatus. All experimental stimuli were drawn from a set of 2,664 colored photographic images of natural scenes and a corresponding set of 2,664 colored texture synthesis masks created using Portilla and Simoncelli's (2000) algorithm. Of the 2,664 photographs, 900 depicted only one of our categories; 864 depicted two of the categories (e.g., a human on a beach), with all possible pairs equally represented; and the remaining 900 did not depict any of the nine categories. The images were obtained primarily from a public image data set hosted by the Computational Visual Cognition Laboratory (n.d.) at the Massachusetts Institute of Technology;

some additional images were selected from other Web and personal archives. The stimuli subtended $13^\circ \times 13^\circ$ of visual angle at the viewing distance of approximately 57.3 cm. Stimuli were presented on a 21-in. monitor (resolution: 1024×768 ; refresh rate: 75 Hz) controlled by a Macintosh G5 computer running Mac OS 10.4. The experiment was controlled by MATLAB 7.5.0 and the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997).

Procedure. Each trial consisted of a rapid serial visual presentation (RSVP) of six images. Following an initial 300-ms fixation, the six images were presented sequentially, for 20 ms each (Fig. 1a). The second of these images was the photographic scene, and the remainder were synthesized texture masks derived from the statistical properties of other scenes in the image set. Each scene was unique.

Observers were asked to indicate whether or not a specified cued category appeared in the RSVP stream. We compared a condition in which the target category was named before the stream of images (precue) with a condition in which it was named after the stream (postcue). Precues were presented for 800 ms before the initial 300-ms fixation preceding the RSVP stream, and postcues were presented for 800 ms before the final response-request display. On each trial, the observer was cued with a target category randomly selected from the nine

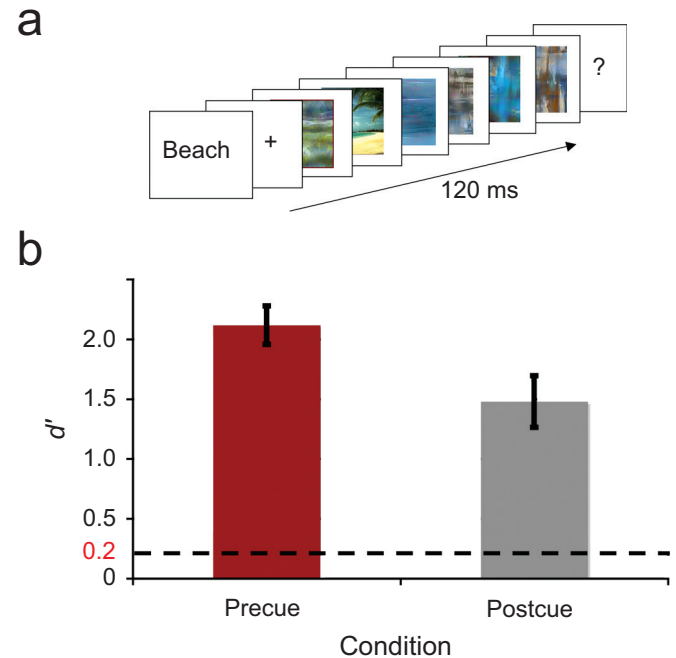


Fig. 1. Stimuli and results from Experiment I. On each trial, a photograph of a natural scene was presented within a stream of synthesized texture masks (a). Observers were asked whether the scene represented one of nine possible categories (animal, human, vehicle, bridge, flower, mountain, beach, street, or indoor scene). The category to look for was indicated either before (precue condition; illustrated here) or after (postcue condition) the stimulus stream. The graph (b) presents the signal detection sensitivity parameter, d' , as a function of condition. The dashed line represents the value of d' at chance. Error bars represent standard errors of the mean.

possible categories (animal, human, vehicle, bridge, flower, mountain, beach, street, or indoor). The target category was present on 50% of the 1,800 trials.

Data analysis. We converted accuracy to d' because d' is theoretically independent of an observer's bias to respond "yes" or "no." To test the hypothesis that observers monitored for only one category at a time, we calculated the expected d' for the postcue condition by assuming that when observers happened to monitor the correct category, their performance would be equal to that observed when the same category was precued, but when they monitored the wrong category, their performance would be at chance. We computed this predicted d' from the observed d' in the precue condition and the observed criterion (c) in the postcue condition according to Equation 1:

$$d'_{\text{predicted}} = z \left[\frac{\Phi\left(\frac{d'_{\text{precue}}}{2} - c_{\text{postcue}}\right) + (n-1)\Phi(-c_{\text{postcue}})}{n} \right] - z \left[\frac{\Phi\left(-c_{\text{postcue}} - \frac{d'_{\text{precue}}}{2}\right) + (n-1)\Phi(-c_{\text{postcue}})}{n} \right] \quad (1)$$

In this equation, z refers to a z-score computation, Φ to the area under the cumulative normal distribution, and n to the number of categories. The logic is as follows: One computes d' by taking the z score of the hit rate and subtracting the z score of the false alarm rate (Macmillan & Creelman, 1991), so in order to compute the predicted d' , one needs to know the predicted hit and false alarm rates under the hypothesized scenario (i.e., observers were monitoring for one category at a time). In order to generate these predicted hit and false alarm rates, we work backward from the observed d' in the precue condition, which estimates the observer's sensitivity when monitoring a single category. Computing the area under the normal distribution (indicated by Φ) at $(d'_{\text{precue}}/2)$ gives the expected hit rate (at a neutral criterion, c) for the category that the observer is monitoring. Of course, the actual hit rate depends on the criterion. The observed criterion for the postcue condition, c_{postcue} , turns out to be a good estimate of the hypothetical criterion when the observer is monitoring a single category, so the expression $\Phi\left(\frac{d'_{\text{precue}}}{2} - c_{\text{postcue}}\right)$ gives the predicted hit rate when the observer happens to be monitoring the correct category. If the observer is monitoring the wrong category, performance will be at chance, giving a d' of 0 and a hit rate of .5 at a neutral criterion. The hit rate is again adjusted according to c_{postcue} . Multiplying

this chance value by $(n - 1)$ and dividing the whole by n weights the chance component by the number of nonmonitored categories. The same logic holds for the computation of the predicted false alarm rate.

Results and discussion

Unsurprisingly, performance was significantly better with a precue than with a postcue, $t(9) = 9.30, p < .01$ (see Fig. 1). The advantage for the precue condition is evident not only from d' , but also from percentage correct; accuracy was 83% ($SEM = 1\%$) in the precue condition and 76% ($SEM = 2\%$) in the postcue condition. More important, performance in the postcue condition was well above chance, $t(9) = 8.69, p < .01$. In the postcue condition, did observers perhaps try to guess which category would be cued, monitor that category (as in the precue condition), and respond at random on trials when they guessed wrong? Under this scenario, Equation 1 predicts a d' of 0.20, which is significantly less than the observed d' of 1.48, $t(9) = 14.36, p < .00001$. Our results therefore imply that observers can monitor for multiple scene categories simultaneously.

Experiments 2 and 3

What happens when more than one potentially relevant category is present in a scene? For example, is an animal on a beach more difficult to recognize than an animal in another setting if "beach" is a cued target on other trials? To answer this question, in Experiments 2 and 3, we added critical trials, which contained more than one potential target category.

Twelve observers (5 females, 7 males; age range = 19–45 years) participated in Experiment 2, and 16 observers (7 females, 9 males; age range = 20–52 years) participated in Experiment 3. The task was the same as in Experiment 1: to indicate whether a single specific pre- or postcued target category was present in a scene (Fig. 2a). Each critical scene contained instances of both the category cued as the target for that trial (the *trial-relevant category*) and another category that was relevant to the task but not cued for that trial (a *task-relevant category*). Thus, for example, in the case of a critical scene of an animal on a beach, if "beach" was the trial-relevant (target) category, "animal" was a task-relevant category (because it was the target category on other trials); alternatively, if "animal" was the trial-relevant (target) category, "beach" was a task-relevant category. There were four trial types defined by the categories present: (a) both a trial-relevant and a task-relevant category present, (b) only a trial-relevant category present, (c) only a task-relevant category present, and (d) no relevant category present. In Experiment 2, exposure duration was fixed at 20 ms (as in Experiment 1), whereas in Experiment 3, we varied exposure durations from 20 to 200 ms. We report accuracy rather than d' , as it is unclear how to compute d' when there are multiple types of target-absent trials (i.e., the last two trial types).

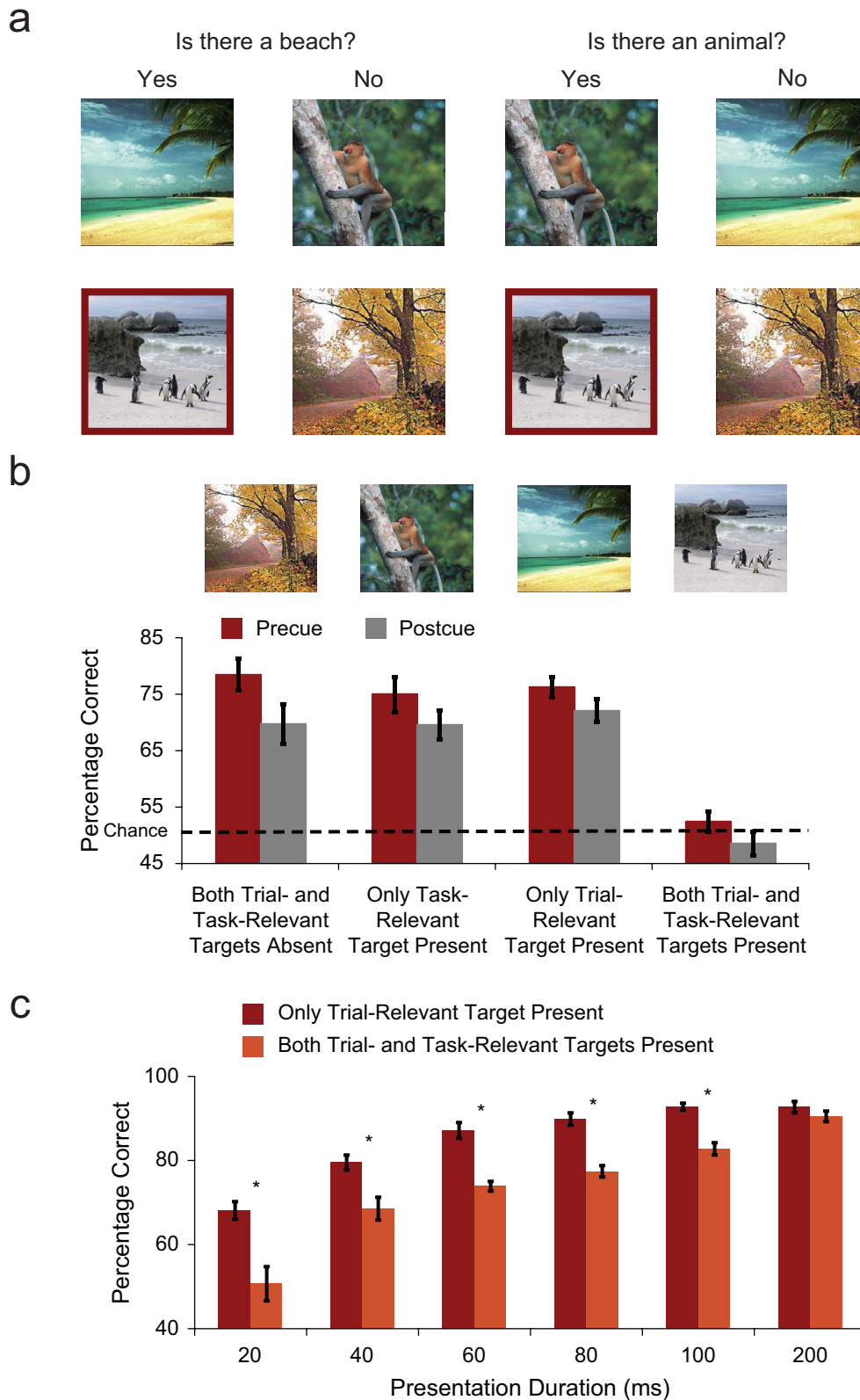


Fig. 2. Stimuli and results from Experiments 2 and 3. The examples in (a) illustrate the four trial types, defined by whether a task-relevant but uncued category (e.g., animal, beach) was present and whether the cued (trial-relevant) category was present. When the scene included images belonging to both the cued, trial-relevant category and the uncued, task-relevant category (examples highlighted by the red frames), one category was cued for some observers (randomly determined), and the other was cued for the other observers. The graphs show (b) mean accuracy as a function of trial type and condition (precue or postcue) in Experiment 2 ($N = 12$) and (c) mean accuracy as a function of trial type and exposure duration in Experiment 3 ($N = 16$). Error bars represent standard errors of the mean. Asterisks indicate a significant difference between trial types ($p < .05$).

The most important finding is that the observers in Experiment 2 were at chance in reporting the presence of the trial-relevant category if the image also contained an uncued task-relevant category (see Fig. 2b for results for the four trial types). This was true not only when the target category was postcued (testing deviation from chance), $t(11) = -0.73, p = .48$, but also when the target category was precued, $t(11) = 1.75, p = .10$. For example, the presence of an animal made observers significantly less likely to successfully report a beach, even when “beach” was cued as the target before the presentation of the scene. All categories interfered with each other equally. In Experiment 3, this destructive interference was observed for all exposure durations less than 200 ms (Fig. 2c). Performance was at chance at the 20-ms exposure and above chance but still impaired at the 40-ms exposure, $t(13) = 4.66, p < .01$. By the 200-ms exposure, there was no interference effect (one category: 93% correct, $SEM = 1\%$; two categories: 90% correct, $SEM = 2\%$), $t(13) = 2.03, p = .07$. Thus, in Experiments 2 and 3, categories collided destructively.

Experiment 4

Perhaps the interference observed in Experiments 2 and 3 was due to the scenes with two categories being more complex than the others. In the next series of experiments, we tested this possibility by manipulating whether the second category in the critical scenes was task relevant or not. In Experiment 4 (20 observers; 10 females, 10 males; age range = 19–36 years), the stimuli were the same as in Experiment 2. In this experiment, only the precue condition was included. There were six

possible target categories. In the first block of trials, only three (randomly chosen for each observer) of these categories were cued and therefore task relevant. In the second block, all six categories were cued and therefore potentially relevant, but during the first block, observers had no way to know that the other three categories would become relevant later. In this experiment, there were six trial types (see Fig. 3): (a) both a trial-relevant and a task-relevant category present, (b) both a trial-relevant and a non-task-relevant category present (Block 1 only), (c) only a trial-relevant category present, (d) only a task-relevant category present, (e), only a non-task-relevant category present (Block 1 only), and (f) none of the six categories present. Consider an image containing a beach and an animal on trials in which “beach” is the trial-relevant target. The critical comparison would be between such a trial in Block 1, when animals were never targets, and in Block 2, when animals were targets on one sixth of the trials.

Percentage correct for the six trial types in Blocks 1 and 2 is presented in Figure 3. The critical result is that performance dropped significantly from 73% (Block 1) to 56% (Block 2) on trials in which the target category was paired with a category that became task relevant in Block 2 only, $t(19) = 12.83, p < .01$. Thus, nontarget categories interfered only when they were task relevant. Data from Block 1 also indicate that interference was induced by task relevance rather than scene complexity. Performance was 54% correct when trial- and task-relevant categories were present in Block 1 and 73% correct when essentially equivalent images were presented but the second category was not currently task relevant (cf. the first two columns in the upper row of Fig. 3). Note that







Nontarget \ Target	Mountain (Relevant in Both Blocks)	Animal (Silent in Block 1, Relevant in Block 2)	No Mountain, No Animal
Beach (Relevant)	 54% 649 ms 57% 674 ms	 73% 602 ms 56% 649 ms	 75% 571 ms 76% 578 ms
No Beach	 70% 705 ms 75% 709 ms	 76% 645 ms 76% 714 ms	 76% 652 ms 79% 648 ms

Fig. 3. Stimuli and results from Experiments 4 and 5. For illustrative purposes, trial types are labeled with reference to “beach” as the cued target category, “mountain” as the task-relevant category, and “animal” as the non-task-relevant category (in Block 1). Target-present trials are represented in the top row, and target-absent trials are represented in the bottom row. The columns represent (in order from left to right) trials with a second category that was task relevant in both Blocks 1 and 2, trials with a second category that was task relevant in Block 2 only, and trials without a second task-relevant category. The numbers at the bottom of the images are percentage correct (Experiment 4) and mean reaction time (Experiment 5); results for Block 1 are printed in black, and results for Block 2 are printed in red.

performance dropped to chance when both trial-relevant and task-relevant categories collided in the same trial. Otherwise, performance was well above chance. Given that the task-relevant and task-irrelevant categories of Block 1 were counter-balanced across observers, this difference in performance must be due to task relevance. Thus, interference between categories did not stem from the visual properties of the scenes or from factors outside the experiment (e.g., some bias against reporting the presence of humans when animals were present).

Experiment 5

In the first four experiments, we used extremely brief presentations. Experiment 3 showed that the interference effect on accuracy was eliminated when the exposure duration was a still fairly brief 200 ms. Perhaps the observed interference was an artifact of the extremely brief exposure durations. In Experiment 5 (12 observers; 9 females, 3 males; age range = 21–53 years), observers had unlimited time to view each scene, and reaction time was the dependent measure. The design was the same as in Experiment 4, except that instead of being presented briefly as part of an RSVP stream, scenes were presented unmasked until the observer responded. We analyzed reaction times from only those trials on which observers responded correctly and eliminated trials with reaction times less than 200 ms or more than 3 standard deviations above the mean.

Figure 3 presents results for the six trial types in Blocks 1 and 2. Reaction time to the trial-relevant category was significantly slower when a task-relevant category was present than when the second, uncued category was not task relevant, $t(11) = 2.69, p < .02$. Therefore, the interference observed in Experiments 2 through 4 was not just an artifact of rapid scene presentation, but also can be seen in reaction times even when the images are not masked and participants have an unlimited time to view them.

Experiment 6

Perhaps the interference effect was due to observers seeing the uncued but task-relevant category at the expense of the cued category in the same image. We tested this hypothesis in Experiment 6 (20 observers; 8 females, 12 males; age range = 19–50 years), using the same design as in Experiment 4. After reporting whether the precued trial-relevant category was present, observers reported if any other categories on a list they were given were also present. In fact, missing the cued target rendered observers *less* likely to successfully report the second, nontarget category, $t(18) = 5.43, p < .01$. When categories collide, the resulting interference is mutually destructive for the trial-relevant and task-relevant categories.

Experiments 7 and 8

In our first six experiments, we asked observers to report on the presence of a single, trial-relevant category. In two final

experiments, we asked observers to report on the presence of two categories at the same time.

Experiments 7 and 8 used the same design as Experiment 2, except that two categories were precued simultaneously (the postcue condition was not included). In Experiment 7 (12 observers; 8 females, 4 males; age range = 21–54 years), the task was to report if both of the precued target categories were present (*and* condition). There were an equal number of trials when the correct answer was “yes” and when it was “no.” In Experiment 8 (12 observers; 5 females, 7 males; age range = 19–45 years), the task was to report if either one of the precued categories was present (*or* condition). Again, there was a 50% chance that the correct answer was “yes.” All possible pairs of the nine different categories were tested in both experiments.

Figure 4 shows percentage correct as a function of the number of trial- or task-relevant categories present in Experiments 2, 7, and 8. Images with no task-relevant categories always required a “no” response. Images with two task-relevant categories always required a “yes.” Images with one target category present required a “yes” in Experiment 8 and a “no” in Experiment 7. In Experiment 2, the one-category trials always showed the target, so “yes” responses were required. Notice that the patterns of interaction in Figure 4 depend on task demands. In Experiment 2, categories collided in the case of two-category images, and performance was reduced to guessing on two-category trials. In the *and* condition (Experiment 7), observers reported “no” accurately when no category was present, but less accurately when one target was present. Performance was not specifically impaired on two-target, “yes” trials. In the *or* condition (Experiment 8), performance was actually facilitated on two-category trials relative to one-category trials (e.g., accuracy was higher for a scene of an animal on a beach than for scenes with just an animal or just a beach), $t(11) = 12.26, p < .0001$. Thus, the presence of a second category can be destructive (Experiments 2–6) or constructive (Experiment 8).

Conclusion

What kind of model can explain this set of results? Suppose that the presence of a category is detected when information accumulates to some threshold. Experiments 1 and 2 show that nonselective information about multiple categories accumulates in parallel, whereas Experiments 2 through 6 show that sometimes information about the presence of one category can be taken as evidence *against* the presence of another. Such use of a category’s presence as evidence against the presence of another category is a function of the probability structure of the experiment. Thus, in a search for a trial-relevant animal on a task-relevant beach, information about the beach may be taken as negative evidence about the animal. If the scene is exposed only briefly (Experiments 2, 4, and 6), “animal” information fails to reach the detection threshold, and performance is at chance. If exposure is unlimited (Experiment 5), it takes longer for this information to reach threshold. When two

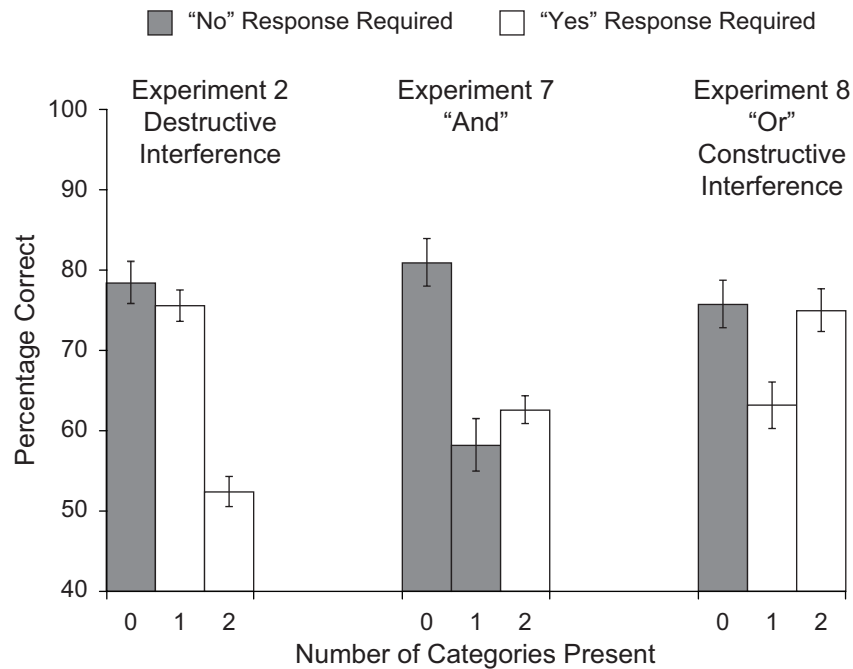


Fig. 4. Results from Experiments 2, 7, and 8. Percentage correct is shown as a function of the number of categories present in the image. In Experiment 2, observers reported whether a single specified category was present (only results for the precise condition are included here); in Experiment 7, they reported whether both of two specified categories were present; and in Experiment 8, they reported whether either of two specified categories was present. Error bars represent standard errors of the mean.

categories are cued (Experiments 7 and 8), however, the presence of one category is no longer taken as negative information about the other. Indeed, probability summation should enhance two-category performance in the *or* condition because a positive response can be generated if information for either category reaches threshold.

In sum, multiple nonselective processes accumulate information about different properties simultaneously and flexibly. The outputs of these processes can interfere destructively or constructively, depending on the task at hand. This finding might point to some of the limitations of making decisions in the proverbial “blink of an eye,” or in the case of “thinking without thinking” (Gladwell, 2005). Although remarkable amounts of information can be extracted from the world in a very brief time, people’s understanding of the contingencies of the world clearly influence how they use that information. If their underlying theories are wrong, they may let one bit of information destroy or accumulate with another in a manner that could lead to an incorrect conclusion. “In the night, imagining some fear, how easy is a bush supposed a bear” (Shakespeare, ca. 1595/1997: *Midsummer’s Night Dream*, Act 5, Scene 1, line 21).

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was funded by Ruth L. Kirschstein National Research Service Award Grant F32EY019819-01 to Karla K. Evans and by National Institutes of Health–National Eye Institute Grant EY17001 and Office of Naval Research Multidisciplinary University Research Initiative Grant N000141010278 to Jeremy M. Wolfe.

References

- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404.
- Chubb, C., Nam, J. H., Bindman, D. R., & Sperling, G. (2007). The three dimensions of human visual sensitivity to first-order contrast statistics. *Vision Research, 47*, 2237–2248.
- Computational Visual Cognition Laboratory. (n.d.). *Urban and natural scene categories*. Retrieved from <http://cvcl.mit.edu/database.htm>
- 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.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), Article 10. Retrieved from <http://www.journalofvision.org/content/7/1/10.full?sid=931b6a17-68c2-4cbe-a2f6-bf19c3ce2dd6>

- Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, *15*, 455–469.
- Gladwell, M. (2005). *Blink: The power of thinking without thinking*. New York, NY: Little, Brown and Co.
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*, 137–176.
- Ishihara, S. (1987). *Test for colour-blindness*. Tokyo, Japan: Kanehara.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203.
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*, 3286–3297.
- Kaplan, S. (1992). Environmental preference in a knowledge-seeking, knowledge-using organism. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 581–600). New York, NY: Oxford University Press.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, USA*, *99*, 9596–9601.
- Macmillan, N. A., & Creelman, C. D. (1991). *Signal detection theory: A user's guide*. Cambridge, England: Cambridge University Press.
- Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research*, *39*, 2929–2946.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–70.
- Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*, *253*, 437–438.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, *3*, 1199–1204.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, *8*, 363–370.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, USA*, *104*, 6424–6429.
- Shakespeare, W. (1997). A midsummer night's dream. In G. Blakemore Evans & J. J. M. Tobin (Eds.), *The Riverside Shakespeare* (pp. 256–283). Boston, MA: Houghton Mifflin. (Original work published ca. 1595)
- Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, *32*, 579–592.
- Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, *14*, 715–725.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, *30*, 655–668.
- Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), Article 8. Retrieved from <http://www.journalofvision.org/content/5/1/8.full?sid=eecd6313-d77c-45a9-b44b-78b426c34651>