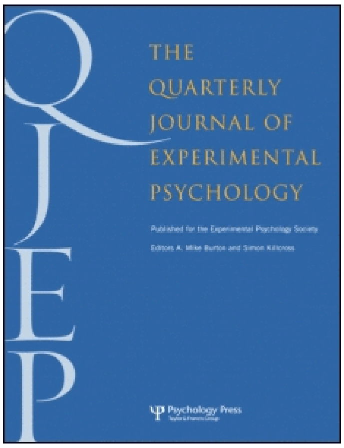

The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

Detecting objects is easier than categorizing them

Jeffrey S. Bowers [a]; Keely W. Jones [a]
[a] University of Bristol, Bristol, UK

First published on: 17 December 2007

To cite this Article Bowers, Jeffrey S. and Jones, Keely W.(2008) 'Detecting objects is easier than categorizing them', The Quarterly Journal of Experimental Psychology, 61: 4, 552 — 557, First published on: 17 December 2007 (iFirst)

To link to this Article: DOI: 10.1080/17470210701798290

URL: http://dx.doi.org/10.1080/17470210701798290

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2008, 61 (4), 552–557

Psychology Press
Taylor & Francis Group

Short article

Detecting objects is easier than categorizing them

Jeffrey S. Bowers and Keely W. Jones
University of Bristol, Bristol, UK

Two experiments compared performance in an object detection task, in which participants categorized photographs as objects and nonobject textures, and an object categorization task, in which photographs were categorized into basic-level categories. The basic-level categorization task was either easy (e.g., *dogs* vs. *buses*) or difficult (e.g., *dogs* vs. *cats*). Participants performed similarly in the detection and the easy-categorization tasks, but response times to the difficult-categorization task were slower. This latter finding is difficult to reconcile with the conclusions of Grill-Spector and Kanwisher (2005) who reported equivalent performance on detection and basic-level categorization tasks and took this as evidence that figure–ground segregation and basic-level categorization are mediated by the same mechanism.

Keywords: Object recognition; Basic level categorization; Object identification; Figure-ground segregation.

The speed with which humans identify objects belies the complexity of the process. According to standard theories, various low-level computations are first performed on an image projected on the retina, including edge extraction (Marr & Nishihara, 1978), depth segregation (Nakayama, Shimojo, & Silverman, 1989), and the detection of nonaccidental properties such as colinearity, curvature, and cotermination (Biederman, 1987). Figure–ground segregation is then computed at an intermediate stage of processing (Driver & Baylis, 1996), followed by basic- and subordinate-level categorizations in the final stages (e.g., Nakayama, He, & Shimojo, 1995). It is widely assumed that early, intermediate, and late stages of visual processing are completed in sequence, even when top-down feedback is in play.

Given this general framework, it is striking that Grill-Spector and Kanwisher (2005) provide evidence that basic-level categorization occurs at the same time (and at the same processing step) as putatively earlier stages of processing, such as figure–ground segregation. In a series of studies the authors assessed the relative amount of time required to perform three tasks designed to tap into different stages of object identification. Colour photographs of objects from various basic-level categories were presented over a range of exposure durations (17, 33, 50, 68, or 167 ms). In a detection task, an image of an

Correspondence should be addressed to Jeffrey S. Bowers, Department of Experimental Psychology, University of Bristol, 12a Priory Road, Clifton, Bristol BS8 1TU, UK. E-mail: j.bowers@bris.ac.uk
We would like to thank Chris Benton for writing a Matlab programme for us that generated the nonobject textures.

http://www.psypress.com/qjep DOI:10.1080/17470210701798290

intact object or a nonobject texture (created by randomly scrambling object images) was presented on each trial, and participants made object/nonobject decisions. Given that the textures were effectively random noise, object responses could be performed on the basis of detecting a line segment, or any low-level feature of an object. In the categorization task, pictured objects from different basic-level categories were presented, and participants categorized them into two categories (e.g., *cats* vs. *other animals*). Responses were assumed to require categorization of the objects at the basic level. Finally, in an identification task, participants made subordinate-level categorizations (e.g., *German Shepherds* vs. *other dogs*).

Not surprisingly, participants were faster and/or more accurate in the basic-level categorization task than in the subordinate identification task, consistent with previous findings (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Surprisingly, however, performance was equivalent in the detection and categorization tasks. That is, participants found it equally easy to distinguish a cat from other animals as between a cat and a random pattern of dots. Based on these findings, the authors concluded: "as soon as you know it is there, you know what it is" (Grill-Spector & Kanwisher, 2005, p. 152). The authors took these findings to challenge standard models of object identification in which figure–ground segregation precedes object categorization. Instead, the authors claimed that a common process supports both functions.

In the present article we challenge these conclusions. The claim that we detect and categorize objects at the same time cannot be true in general, as we do know something is moving in peripheral vision before we know what it is. However, we also question whether it is true in the context of identifying static images. Two features of Grill-Spector and Kanwisher's (2005) studies may have led to a false conclusion. First, in three of their four experiments, detection and categorization accuracy were assessed for briefly flashed images immediately followed by a pattern mask. The authors took the equivalent performance in these tasks as evidence that detection and categorization take the same amount of time. The problem with this approach (as the authors note themselves) is that masking may be more effective in blocking from awareness the outputs of low-level than high-level vision, and as a consequence, participants may have been forced to rely on the same relatively high-level processes in both tasks. This would render any general inferences about the relative time-course of detection and categorization invalid. In the same way, word identification is better than letter identification under masking conditions (the word superiority effect), but this is not taken to reflect the relative time-course of identifying letters and words. That is, no one takes the word superiority effect to indicate that word identification precedes letter identification. The problem with the masking procedure may be particularly acute in the present case, as the nonobject textures and masks were visually similar to one another (see Figure 1 from Grill-Spector & Kanwisher, 2005), which might artificially impair performance in the detection task.

This problem was addressed in Grill-Spector and Kanwisher's (2005) Experiment 3 in which detection and categorization were compared without masking. Images were again presented for a range of durations, and again performance in the two tasks was the same, supporting their conclusion. However, in this study, the categorization task was made relatively easy by asking participants to distinguish between objects from different superordinate categories (e.g., *cars* vs. *objects*). Under these conditions, participants may categorize objects based on detecting diagnostic low-level visual features (cf. Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005); that is, these simple categorizations might not require basic-level categorizations at all.

We report two experiments comparing performance on a detection and a basic-level categorization task without masking when the basic-level categorizations were relatively easy (when objects were from different superordinate categories; e.g., *trains* vs. *dogs*) and when categorizations were more difficult (when objects were from the same superordinate category; e.g., *dogs* vs. *cats*). The easy condition is similar to Grill-Spector and Kanwisher's (2005) Experiment 3, and the difficult condition was included to rule out the hypothesis that low-level visual features could support categorization performance. Note, our difficult-categorization task differs from Grill-Spector and Kanwisher's identification task (e.g., *German Shepherds* vs. *other dogs*), as the former constitutes a basic-level categorization task, and the latter constitutes a subordinate-level categorization task. The hypothesis we are concerned with here is whether basic-level categorization occurs at the same time as object detection.

Pictures were presented for 500 ms in the first study and 50 ms in the second. This second study was included because the 50-ms duration matches one of the exposure conditions of Grill-Spector and Kanwisher's (2005) Experiment 3. But in both of our studies the images were clearly presented, and the critical dependent measure is reaction time (RT). If indeed detection and basic-level categorization co-occur, RTs should be the same in the detection, easy-categorization, and difficult-categorization conditions.

Method

Participants

A total of 16 participants completed each study. All had normal or corrected-to-normal vision (by self-report).

Materials and design

A total of 480 images were chosen, 32 for each basic-level category, plus a further 32 images for a "multiple-object condition" block used in the detection task (described below). Seven pairs of basic-level categories were used, organized into six superordinate categories (animals: cats/dogs and horses/cows; fruits: apples/pears; flowers: daffodils/sunflowers; vehicles: trains/buses; tools: hammers/wrenches; musical instruments: guitars/violins). Colour photographs from each basic-level category were retrieved from the Internet and were resized to equal dimensions of 375 × 375 pixels. A pilot study ensured that the objects were easily identifiable. Nonobject textures were created by randomly scrambling each pixel of each picture. In this way the object/nonobject images were matched for colour and luminance. The experiment was run using the DMDX presentation software (Forster & Forster, 2003).

The experiment included task as a within-participant factor with three levels: (a) detection, in which objects were intermixed with nonobject textures, and participants pressed one shift key in response to objects and the other shift key in response to nonobjects; (b) easy categorization, in which objects from two basic-level categories from different superordinate level categories were intermixed, and participants pressed one shift key in response to one category and the other shift key in response to the other category; (c) a difficult categorization in which objects from two basic-level categories from the same superordinate level category were intermixed, and participants pressed one or the other shift key in response to the two basic-level categories. The detection task itself included two conditions. Either nonobject textures were intermixed with objects from one category (e.g., *nonobject textures* vs. *dogs*; the "single-object condition"), or nonobject textures were intermixed with objects from all the categories (e.g., *nonobject textures* vs. *dog*s or *buses* or *violins* etc.; the "multiple-object condition"). The latter condition was included to ensure that detection was not based on perceiving a simple visual feature associated with all the objects in a given category. In all task conditions the assignment of left/right responses was counterbalanced across participants. Participants completed three blocks of trials in each of the task conditions, with different pictures in each block, and blocks were presented in eight different randomized orders, counterbalanced across participants. In the detection task, participants completed two blocks in the single-object condition and one block in the multiple-object condition. The presentation order of images within each block was randomized. An equal number of exemplars from the two categories was presented within each block, and no picture was repeated. In addition, two different assignments of pictures

Table 1. *Mean reaction times, percentages of errors, and standard deviations for each task in response to objects and textures in Experiments 1 and 2*

	RT (ms)	*SD*	*Errors (%)*	*SD*
Experiment 1				
Overall detection (objects & textures)	427	41	4.7	2.1
Easy categorization	447	41	5.2	2.9
Difficult categorization	507	46	7.4	3.1
Object detection (single-object condition)	425	46	3.8	2.6
Object detection (multiple-object condition)	472	59	7.8	5.6
Overall object detection (single- & multiple-object conditions)	441	48	5.1	2.6
Experiment 2				
Overall detection (objects & textures)	401	57	5.9	3.9
Easy categorization	425	58	6.8	4.2
Difficult categorization	488	66	10.7	7.1
Object detection (single-object condition)	397	60	6.4	4.2
Object detection (multiple-object condition)	441	72	6.3	5.1
Overall object detection (single- & multiple-object conditions)	411	63	6.4	3.4

to the different task conditions were included. For example, for half of the participants *apples* were assigned to a detection block, and for the other participants, *apples* were assigned to a difficult-categorization block. This was intended to reduce the likelihood of any stimulus-specific effects on task performance.

Procedure

On each trial an image was presented for 500 ms and 50 ms in Experiments 1 and 2, respectively, and was immediately replaced by a neutral grey screen, which remained until a response was made, or for a maximum of 4,000 ms. Each block started with instructions indicating which categorizations were to follow and consisted of 6 practice and 64 experimental trials (e.g., 32 *dogs* and 32 *cats*), so that there were 192 critical trials in each task, 576 in total. Participants categorized the pictures according to the task condition as quickly as possible by pressing the left and right shift keys, and feedback was given after every trial.

Results

RTs greater than 1,000 ms, those less than 200 ms, and errors were all excluded from analysis of the RTs (5.8% of trials in Experiment 1, 7.8% of trials in Experiment 2). Mean RTs and errors for the two experiments are shown in Table 1. In both experiments, overall detection RTs (collapsing across the objects and textures) were significantly faster than the easy categorizations, $t(15)$s > 4.1, p values $<.01$, which were in turn faster than the difficult categorizations, $t(15)$s >100, p values $<.01$. Indeed, collapsing across experiments, 28/32 participants were faster in the overall detection than in easy categorization, and 32/32 were faster in the easy than in the difficult-categorization task. The error results were in the same direction, so there was no speed–accuracy trade-off. Thus, contrary to Grill-Spector and Kanwisher (2005), these results suggest that detection precedes categorization when masking is not a factor.

One possible problem with this analysis is that we have collapsed across all images in all tasks. This makes sense when analysing the categorization responses, as all the images were objects. But in the case of the detection task, we have collapsed across responses to objects and nonobject textures. It is possible that RTs to textures were faster than those to objects, and a different conclusion would have been reached if we only analysed responses to objects. Furthermore, we have collapsed across the single- and multiple-object conditions in the detection task, and the composition of objects in the detection task may impact on the RTs.

In Table 1 we also report mean RT and error rates for the detection of objects (excluding textures) in the single- and multiple-object conditions, as well as overall object detection (collapsing across single- and multiple-object conditions). In fact, overall object detection times are similar to the easy-categorization times. In Experiment 1 overall object detection RTs (441 ms) and easy-categorization RTs (447 ms) did not differ, $t(15) = 1.37$, $p > .15$, whereas in Experiment 2 overall object detection times (411 ms) were still faster than easy-categorization times (425 ms), $t(15) = 3.13$, $p < .01$. The error rates in the two conditions were also very similar. Thus although we still obtain some evidence that detection precedes categorization when only analysing the responses to the objects, the results are not so clear-cut. In addition, the composition of objects in the detection task had a large effect, with object detection in the multiple-object condition slower than that in the easy-categorization condition in both experiments. Thus the conclusion that one makes regarding the relative difficulty of the detection and easy-categorization tasks depends on which specific detection conditions are considered. It is not immediately clear which is the appropriate comparison, and, accordingly, it is difficult to reject the hypothesis that detection and easy categorizations occur at the same time based on these data.

Although the above findings can be interpreted as consistent with Grill-Spector and Kanwisher's (2005) findings, the important point to note is that the easy categorizations might not actually require the objects to be categorized at a basic level; detection of a low-level visual feature associated with a category might suffice, in which case, the easy-categorization task is not functionally different from the detection task. Thus, the critical finding is that no matter what criterion one adopts regarding detection, RTs in the difficult-categorization task are substantially slower. Here participants need to categorize the objects at a basic level, as low-level features in the image should be relatively undiagnostic regarding category membership. In Experiment 1, object detection in the multiple-object condition was 35 ms faster than responses in the difficult-categorization condition, $t(15) = 3.42$, $p < .01$, with 13/16 participants slower in the categorization task, and in Experiment 2 the difference was 47 ms, $t(15) = 5.65$, $p < .01$, with 15/16 participants showing the effect. As can be seen from Table 1, there is no evidence for a speed–accuracy trade-off.

GENERAL DISCUSSION

Grill-Spector and Kanwisher (2005) challenged a standard assumption regarding the processing steps involved in object identification based on their finding that object detection and object categorization take the same amount of time to complete. In particular, the similar RTs in the two tasks were taken to challenge the common view that figure–ground segregation precedes object categorization (e.g., Driver & Baylis, 1996; Nakayama et al., 1995). A number of authors have argued that categorization influences segmentation (e.g., Peterson & Gibson, 1993), but based on their RT results, Grill-Spector and Kanwisher (2005) concluded that object segmentation and categorization are based on the same mechanism.

However, the current study provides clear-cut evidence that it takes less time to detect than categorize an object when visual processing is not affected by masking and when the basic-level objects are from the same superordinate category (e.g., *dogs* vs. *cats*). As noted earlier, the masking condition employed by Grill-Spector and Kanwisher (2005) may have effectively blocked awareness of low-level visual properties of objects, which would render any general inferences about the relative time-course of detection and categorization invalid. Similarly, their use of a categorization task in which objects were from different superordinate categories (e.g., *cars* vs. *objects*) might allow participants to categorize objects on the basis of detecting diagnostic visual features of a category, in which case the task is effectively another detection task. Indeed, we also found similar detection and categorization RTs when the to-be-categorized objects were from different superordinate categories. When these methodological problems were addressed by including a difficult basic-level categorization task in which all the objects were from the same superordinate category, these null findings were not replicated, and a robust advantage for detection was observed. We take these findings to be consistent with standard models of object identification in which low-level visual analysis (e.g., edge extraction, etc.) and figure–ground processes precede basic-level categorization.

It should also be noted that Mack, Gauthier, Sadr, and Palmeri (in press) recently reported data problematic for Grill-Spector and Kanwisher's (2005) hypothesis. They carried out two experiments that closely replicated the masked priming conditions of the original study. However, in addition to the standard detection and categorization conditions the authors included a condition in which the objects were inverted or degraded. The logic was that if detection and categorization are mediated by the same processes, then the inverted and degraded conditions should delay both the categorization and detection of the stimuli. By contrast, if different mechanisms are responsible for these two processes, then performance on the two tasks might dissociate. Indeed, this might be expected, given that the features required to distinguish objects from nonobject textures may still be salient in the inverted and degraded conditions.

Consistent with Grill-Spector and Kanwisher (2005), Mack et al. (in press) found equivalent performance in the detection and categorization tasks when the objects were presented in a nondegraded format (and masked). However, detection was superior to categorization performance in the degraded conditions, highlighting the fact that detection and categorization performance can be dissociated. The current findings extend these results by showing that detection precedes categorization under more standard viewing conditions (when images are upright, nondegraded, and unmasked). Taken together, these and the current findings provide a strong falsification of the claims made by Grill-Spector and Kanwisher (2005).

Original manuscript received 27 July 2007
Accepted revision received 29 October 2007
First published online 17 December 2007

REFERENCES

Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure–ground segmentation in short-term visual matching. *Cognitive Psychology*, *31*, 248–306.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research, Methods, Instruments, & Computers*, *35*, 116–124.

Grill-Spector, G., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*, 152–160.

Mack, M., Gauthier, I., Sadr, J., & Palmeri, T. J. (in press). Object detection and basic level categorisation: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*.

Marr, D., & Nishihara, H. K. (1978). Representations of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, Series B*, *200*, 269–294.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *An invitation to cognitive science: Visual cognition* (pp. 1–70). Cambridge, MA: MIT Press.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, *18*, 55–68.

Peterson, M. A., & Gibson, B. S. (1993). Shape recognition contributions to figure–ground organization in three-dimensional displays. *Cognitive Psychology*, *25*, 383–429.

Rosch, E., Mervis, C. B., Gray, M. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.