Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects

Sarah Walker  
Department of Experimental Psychology, University of Cambridge, Cambridge, UK

Paul Stafford  
Department of Experimental Psychology, University of Cambridge, Cambridge, UK

Greg Davis  
Department of Experimental Psychology, University of Cambridge, Cambridge, UK

Human observers can determine whether natural scenes contain an animal or not on the basis of as little as 20 ms viewing; a phenomenon termed ultra-rapid categorization (URC). Recent studies have suggested that URC is unimpaired even when attention resources are concurrently devoted to a second task. This apparent independence of URC from availability of attention resources presents a challenge for the conventional view of high-level vision as attention-demanding. However, one notable feature of the scenes employed in those experiments is that they almost universally comprised only one or two foreground objects. Here, we investigate whether these findings generalize to more complex scenes, more typical of those in nature. We find that categorization of scenes with four primary foreground objects is greatly impaired when attention resources are limited under dual-task conditions, even when scenes are presented for 500 ms. In contrast, URC of scenes with one foreground object is only mildly impaired—the magnitude of this impairment being equivalent to that observed for single objects presented in isolation without naturalistic scene backgrounds. We conclude that URC of complex scenes is particularly attention-dependent but that some attention resources are probably necessary even for URC of simple one-object scenes.

Keywords: gist, scene perception, ultra-rapid categorization, attention


Introduction

Recent behavioral and physiological studies have challenged the prevalent view that high-level visual processing is slow, effortful, and attention demanding. In particular, processing of faces (e.g., Eimer & Holmes, 2002; Liu, Harris, & Kanwisher, 2002), the “gists” of semantic scenes (Biederman, 1972; Biederman, Rabinowitz, Glass, & Stacy, 1974; Potter, 1975, 1976), and detection of the presence of animals (or objects of other well-defined semantic categories) seem to take place rapidly and efficiently. Perhaps the most striking of these demonstrations have addressed the ability of participants to detect animals exceptionally rapidly within natural scenes, termed “ultra-rapid categorization” (e.g., Thorpe, Fize, & Marlot, 1996). Participants are typically able to categorize novel, “natural” scenes as comprising an animal or not on the basis of only 20–30 ms viewing, with concurrent frontal-lobe event-related potentials (ERPs) beginning to diverge in a category-dependent manner within 150 ms (Thorpe et al., 1996), indicating the completion of processing necessary for this categorization within an exceptionally short interval. Earlier studies concluded that such high-level processing was limited to evolutionarily important categories such as predators and food. However, URC has since been shown to occur also for other well-defined non-natural categories such as vehicles (VanRullen & Thorpe, 2001) and therefore potentially provides a more general challenge to conventional views of high-level vision (see e.g., Braun, 2003).

This challenge is most evident in claims that categorization of natural scenes (as opposed to abstract stimulus arrays employed by e.g., Chun & Nakayama, 2000) can arise relatively independently of visual attention and visual short-term memory (VSTM). Li, VanRullen, Koch, and Perona (2002, 2005) have examined URC performance when the task is performed alone, such that participants can devote all of their attention resources to the task, versus when the URC task was performed simultaneously with an attention-demanding second task (searching for an “L” among “T”s at various orientations). Performance on the URC task did not differ significantly between the two conditions, indicating that URC did not depend on the attention resources demanded by the secondary task. Moreover, performance on other non-URC tasks using simple, 2-D, abstract stimuli was significantly affected, indicating that the secondary task did demand significant attention.
attention resources. Together these findings suggest that processing of naturalistic stimuli may show greater independence from attention than 2-D abstract stimuli.

This conclusion has received indirect support from the findings of another recent study by Makovski, Shim, and Jiang (2006), who also showed that performance of a URC task using natural scenes did not disrupt VSTM for displays of four items or fewer. However, this support was only partial. In VSTM displays comprising 8 objects, greatly exceeding the capacity of VSTM (3–4 simple objects; see e.g. Luck & Vogel, 1997), Makovski et al. did find interference from the URC task. The reasons for these discrepant results are unclear. One potential explanation is that performance of the URC task did not affect VSTM in those studies but rather only affected participants’ strategies for dealing with the overloading of their VSTM capacity. Alternatively, it might be that Makovski’s four-object displays had failed to load VSTM capacity sufficiently, leaving spare capacity to perform the URC task.

A further complication in interpreting these results has been revealed by Evans and Treisman (2005), who found that while detecting an animal (or other category of object) demands little in the way of attention resources, locating or more precise identification of the animal does demand limited-capacity attention resources. Evans and Treisman concluded that categorization proceeds on the basis of the presence of many unbound local, low-level features in a scene (of the types thought to be processed independently of visual attention) and need not require sophisticated yet rapid processing of natural scenes. Note, however, that even this rival explanation does still accept that URC can arise relatively independently of visual attention. Our new findings provide the first direct challenge to this assumption.

We have noted one highly salient feature of the natural scenes employed throughout the URC literature that might account for participants’ unexpectedly high performance even in the near-absence of attention. In contrast to the abstract arrays used in visual-search paradigms, scenes in URC paradigms almost universally contain only one or two major foreground objects of which one was always the animal (or comprised many examples of the same animal), if an animal was present. Accordingly, when participants are asked to categorize these scenes as comprising an animal or not, simple salience-based visual mechanisms may guide their attention to the candidate object(s) such that they may only have to recognize one (sometimes two) object(s) within a scene to perform the URC task.

If participants in the URC experiments were indeed only processing one or two easily located objects in order to categorize the scenes, then it is unlikely that URC would present any challenge to conventional views of high-level vision. Alternatively, if participants are able to categorize scenes even when multiple objects are present, URC constitutes one of the greatest challenges to our understanding of high-level vision. We therefore intended to provide a crucial new test to distinguish between these possibilities, by examining URC in scenes with multiple major foreground objects and the role of attention in this process.

One type of capacity limit has already been demonstrated for URC. Rousselet, Thorpe, and Fabre-Thorpe (2004) found that four scenes presented simultaneously could not be processed as efficiently as when only one or two scenes were presented at a time. The amplitude and the latency of URC-associated ERPs was also reduced under the former conditions. However, Rousselet and colleagues’ findings were ambiguous with regard to whether the capacity limitation only arose when multiple scenes were to be categorized simultaneously (a scenario that the visual system will not face in natural environments) or would also have arisen even for a single, sufficiently complex scene. This latter case would be far more typical of perception in everyday life, and accordingly we chose to examine whether a single, complex scene could be categorized “ultra-rapidly” even in the absence of attention resources.

To provide a fair test of whether categorization of multiple foreground object scenes can arise irrespective of the availability of attention and memory resources, the natural scenes in our studies were each presented for either 500 or 170 ms within a trial. This duration permits visual access to the scene (in terms of the scene being physically present) for 6 to 18 times as long as was the case in the Li et al. (2002, 2005) study that concluded URC could arise in the “near absence of attention.”

General method

Participants

Participants for all experiments were recruited by advertisement, aged between 18 and 35 and reported normal or corrected-to-normal vision.

Apparatus

All experiments employed a G4 Apple eMac computer with a 17-in. screen (refresh rate 112 Hz) running PsyScope Software. Default screen luminance and hue were set as neutral gray (50 cdm⁻²). Unless otherwise stated, all images were centralized, with full screen images 800 × 600 pixels, and the maximum angles subtended were 33° horizontally by 25° degrees vertically.

Experiment 1

The logic of Experiment 1 was as follows. The inclusion of a visual short-term memory (VSTM) task in
each trial presumably reduced the amount of VSTM and attention resources available for the secondary task of categorizing scenes as comprising an animal or not (see Fougnie & Marois, 2006; Todd, Fougnie, & Marois, 2005). In one condition (“high load”), three items needed to be encoded and maintained in VSTM and in the other condition (“low load”) only one item; in the former condition, more attention and memory resources should be employed than in the latter, thereby ensuring that fewer resources should be available for scene categorization in the former condition than in the latter. If, as claimed by Li et al. (2002), URC can arise equally efficiently regardless of the availability of attention and memory resources, even in the “virtual absence of attention,” then performance on this task should not differ between the low- and high-load conditions. Conversely, if URC does require limited-capacity attention and memory resources, then performance on the animal-categorization task should suffer in the high- relative to the low-load condition.

**Method**

**Participants**

Nine participants (four male, five female) were tested on both the high- and low-load conditions in separate blocks of 48 trials each. The order was counterbalanced across participants.

**Stimuli, apparatus, and procedure**

Figure 1 schematizes the sequence of displays in a typical high-load trial. Two separate VSTM displays were presented before and after each to-be-categorized scene. These each consisted of either one-colored square (on low-load trials) or three-colored squares (on high-load trials) situated in three possible positions near the center of the screen (see Figure 1, Frame 2); each square subtended approximately 2.3° visual angle. For high-load trials, the squares were each a different colour, with each trial consisting of three of four possible colors: red, green, yellow and blue (indicated by the different gray levels in the figure). On half of the trials, unpredictably, the second display was identical to the first and on the other half, two squares “swapped” colors. For low-load trials, only one square was presented randomly in one of the three positions, in which half of trials were of the same color in the second display as the first and the other half of the trials a different color. Participants responded to the VSTM task using the “+” key to indicate a change between the first and second displays and the “−” key to indicate no change. This response was unspeeded, and auditory feedback signaled an incorrect response.

During the interval between the first and the second VSTM displays, participants were presented with a naturalistic scene and noted whether it comprised an animal or not. In order to compare the first and the second VSTM display, we assume that participants must have attended to and stored the squares in VSTM during the interval between the displays (to anticipate, our findings strongly supported this assumption). Following their response to the VSTM task, the participant verbally indicated the presence/absence of an animal in the scene to the experimenter who pressed one of two keys accordingly. The experimenter could not see the scenes at any point. Each scene was displayed for 500 ms to ensure ample time for processing it.

Each naturalistic scene comprised one of four backgrounds and four randomly selected primary foreground objects (four objects are widely held to be the upper limit of VSTM capacity, e.g., Luck & Vogel, 1997) appearing in four predetermined positions. For each position, there were three possible candidate objects (one animal, two inanimate) to prevent participants deriving the correct response from location alone; this yielded 96 scenes in total (four of which are illustrated in Figure 2). Half of the scenes contained an animal among their candidate objects, and half contained no animal. Since color had been shown to have no significant effect on scene categorization (Delorme, Richard, & Fabre-Thorpe, 2000; Li et al., 2005), all stimuli were monochromatic, with size and positions of objects varying between backgrounds.

Our scenes differed from those in other URC experiments in that individual components were repeated within the same experimental session in order to control for locations and identities of objects. One potential criticism that might be leveled against this approach is that subjects did not see a novel scene on every trial, and so their coding of later scenes may have been affected by regularities across previously viewed scenes. This is indeed a possibility, although we suspect that because each scene (in terms of a particular combination of objects and background) was presented just once in Experiment 1
The benefits of our own approach to designing scenes were that firstly (in contrast to previous URC studies), the background could (demonstrably) provide no reliable information about whether an animal was present at a particular location. This was because the same background was equally likely on any given trial to be associated with an animal in any of the four potential locations for animals in that scene. Secondly, by constructing our scenes from specifically placed objects and backgrounds, we able to create more comparable one-object and four-object scenes in our third experiment.

Taken as a whole, therefore, we considered that this assessment of the relative benefits of the two approaches favored the method we adopted here, rather than that used by previous URC studies.

Results

Accuracy in the memory tasks averaged 92.1% ($\sigma = 3.86$) for the low-load memory task and 75.2% ($\sigma = 6.62$) for the high-load memory task. However, of primary interest were the relative extents to which the low and high VSTM loads had affected performance on the scene-categorization task. So as to ensure that a participant’s response to having made a VSTM-task error on a particular trial did not affect our measure of URC performance (which would have biased our comparisons in favor of low-load trials where fewer VSTM-task errors were made), we considered only those trials in which a correct response was provided to the VSTM task. Even using this conservative index, URC accuracy was significantly poorer in the high-load condition (78.4%, $\sigma = 12.54$) than in the low-load condition (90.2%, $\sigma = 5.41$; $t(8) = 3.522, p < 0.01$).

Discussion

The extra VSTM and attention resources required by the high-load condition relative to the low-load condition of the VSTM task caused a modest but reliable decrease in performance on the animal-categorization task. We concluded that when VSTM capacity is engaged by a second task, URC of scenes is impoverished. Note that the decrease in URC performance arose despite the following two characteristics of the study. First, as indicated above, we presented each scene in the URC task for 500 ms rather than 27 ms as employed by Li and colleagues. This gave ample time for scene categorization, and would normally yield response times of around 450 ms or below. Second, even in our “high-load” task, participants could in fact have responded without error simply by attending any two of the three items in the VSTM display. As two of the items “swapped” colors, at least one of the two attended squares would always have to change whenever a change was present. Accordingly, the high-load task may not even have completely exhausted VSTM capacity, often estimated to be 3–4 simple items (Luck & Vogel, 1997).

Two limitations of this first study, however, restricted our conclusions concerning the availability of attention resources and URC performance. First, because the high- and low-load tasks in Experiment 1 engaged both attention and VSTM resources, it may be that URC performance was only affected when a VSTM load was present and might not have been affected by engaging attention resources alone. To resolve this issue, Experiment 2A employed a URC task in combination with a task that designed to maximize loading of attention resources while minimizing VSTM load. The results of this study were then to be compared to a parallel, complimentary experiment conducted in the same testing sessions (run order counter-balanced across participants) that employed displays similar to those in Experiment 2A but that did involve a substantial VSTM component. This additional study, Experiment 2B, was included to ensure that if we found no effect of loading attention resources alone in

![Figure 2. Sample images of the natural scenes. Each scene consisted of one of four backgrounds (farmyard, garden, beach, and savannah) and 4 candidate objects. On half of the trials, one of these objects was an animal. The animal category included mammals and birds. All scenes were achromatic.](image-url)
Experiment 2A then the differences between the results in this new study, and those of Experiment 1 could not simply be explained in terms of gross differences in the type of stimuli employed in the new task. However, to anticipate our conclusions, this contingency did not arise; Experiment 2A yielded very similar findings to those of Experiment 1B.

If, in these two new studies, we were to find only an effect of manipulating load in Experiment 2B (VSTM load) but not Experiment 2A (Attention load), this would strongly suggest that URC performance was sensitive to allocation of VSTM resources, but not attention resources. Conversely, if we were to find effects of load in both tasks, this would point clearly to the need for focused attention in performing the URC task.

The second limitation inherent in the design of Experiment 1 was that the high- and low-load conditions differed in terms of the physical properties of their VSTM displays and that this difference might have resulted in the drop in URC performance under high-load conditions rather than the manipulation of VSTM load per se. We considered this unlikely, but it in order to preclude such a possibility in Experiments 2A and 2B, we now compared URC performance under dual-task conditions (where a attention or VSTM load was induced by the secondary task) to baseline conditions in which exactly the same sequences of displays were presented but in which participants performed the URC task alone.

**Experiment 2**

In total, URC performance was examined under four conditions. In Experiment 2A, a URC task was performed either alone or simultaneously with another task designed to load attention resources while minimizing VSTM requirements. In Experiment 2B, the same URC task was similarly performed either alone or simultaneously with a VSTM task (similar to Experiment 1 but employing colored letter stimuli identical to those in Experiment 2A). In each case, we compared URC performance under dual-task conditions (performed simultaneously with another task) to URC performance under single-task conditions.

**Method**

**Participants**

Eight new participants (five male, three female) were tested on both Experiments 2A and 2B, with the order counterbalanced across participants. Within each experiment, the run order of single- versus dual-task blocks of trials was similarly counterbalanced across participants. Each block contained 96 trials.

**Experiment 2A—Attention task**

**Stimuli, apparatus, and procedure**

In the attention-loading task performed in Experiment 2A, participants were required to identify whether any one of 4 letters, superimposed over a scene, was a vowel. Letters were presented in the four corners of an 8-cm² square about the center of the scene, and half of the trials consisted of three consonants and one vowel and the other half consisted four consonants. The natural scenes were identical to those in Experiment 1 and contained four major foreground objects. The task was performed under the following conditions: In the single-task condition, participants responded only to the URC task. In the dual-task condition, participants were required to respond to the URC task when a correct response was recorded to the letter task, participants then responded to the URC task, as in Experiment 1. Responses to this secondary categorization task were made contingent upon correct responses in the primary task to discourage strategies of ignoring the primary task in favor of the categorization. In all trials, once both letters and scene had disappeared, the 4-letter positions were masked with a stimulus created from multiple overlaid black letters on a salient white background.

Figure 3A illustrates a typical sequence of displays in a dual-task trial of Experiment 2A. Following a fixation cross (1200 ms), four letters were briefly presented over a background scene containing four primary foreground objects (1200 ms), then masking stimuli in each of the four letter positions, on a gray background (50 ms), gray screen until the next trial. In the dual task condition, participants responded to the attention task using the “+” (vowel present) and “−” (no vowel) keys. Following an incorrect response, auditory feedback was given, and the next trial began immediately. Following correct responses to the attention task, the participants reported verbally to the experimenter (who could not see the stimuli and who entered their response using one of two keys on the keyboard) the presence/absence of an animal in the scene. In single-task trials, participants were shown identical displays but only gave the (verbal) response to the URC task.

**Experiment 2B—Memory task**

**Stimuli, apparatus, and procedure**

The stimuli, apparatus, and procedure of Experiment 2B were similar to those in Experiment 1 but employed colored letter displays analogous to those in the attention task, again employing the single and the dual-task conditions as described above. To remain as comparable to the attention task as possible, the same arrangements of 4 letters were used (taken arbitrarily from the non-vowel list). On half of the trials, the 1st and 2nd letter displays were identical; in the other half, two letters
switched positions between displays. The color assigned to each position remained constant to avoid identification of a change by color alone.

The sequence of displays for a typical dual task trial is schematized in Figure 3B. The sequence consisted of a fixation cross (1200 ms); followed by the first presentation of four-colored letters (to be stored in VSTM) over a background scene containing four major foreground objects. Once the letters had disappeared after 125 ms, the background scene inclusive of the four major objects remained present for another 250 ms. Four-colored letters were then presented for a second time with the background still present (again for 125 ms). The second set of letters were either identical in all respects to the first or two of the letters had exchanged positions. In dual-task trials, participants indicated whether the letters had changed in the second presentation relative to the first by pressing the “+” (change) or “−” (no change) keys. Incorrect responses elicited auditory feedback, and the trial ended. However, following a correct response, participants indicated verbally the presence/absence of an animal in the presented scene. In single-task trials only the verbal response to the URC task was given.

Results

Attention task

Accuracy on the attention (vowel-detection) task averaged only 64.8% (σ = 10.44), indicating that it had been extremely demanding. We then compared URC performance on dual-task trials where a correct response to the attention (vowel-detection) task had been given to that on single-task (URC task only) trials, graphed in Figure 4. Performance in these two conditions averaged 72.0% (σ = 7.32) and 92.4% (σ = 3.92), respectively, a significant drop in URC performance under dual-task relative to single-task trials (t(7) = 6.49, p < 0.01).

Memory task

Participants found the VSTM task less challenging than the attention task in Experiment 2A, averaging 81.3% (σ = 9.23) accuracy. We then compared URC performance under dual-task versus single-task conditions, as shown in Figure 4. Accuracy scores for these two conditions averaged 84.9% (σ = 9.88) and 97.5% (σ = 2.78), respectively, being significantly impoverished under dual-task relative to single-task conditions (t(7) = 3.626, Figure 4).
Quantitative comparison of our findings from Experiments 2A versus 2B is extremely difficult as the displays were physically different (the naturalistic scenes were presented for 170 ms in Experiment 2A, 500 ms in Experiment 2B) and participants clearly found the attention task in Experiment 2A more difficult than the VSTM task in Experiment 2B. Qualitatively, however, it is clear that the same results emerged in each case. URC performance suffered substantially when a second task (that demanded either attention or attention plus VSTM resources) was performed simultaneously with the URC task.

These findings strongly support the conclusion that loading of attention resources impoverishes URC performance, even when memory loads are minimized. They leave unanswered the question of whether a VSTM load alone, in the absence of an attention load, would have the same effect (indeed, it is uncertain whether a VSTM load can ever be induced in the absence of an attention load). One further issue resolved by Experiments 2A and Experiment 2B concerned whether VSTM or attention loads can decrease URC performance even in the absence of potential confounds relating to the nature of the stimuli employed. Both experiments clearly demonstrated an effect of load that could not be attributed to physical aspects of the displays as both dual-task conditions, where an extra VSTM/attention load was present and single-task conditions, in which an extra VSTM/attention load was absent, employed the same stimuli. Our current findings therefore fail to generalize those of Li et al. (2002), who found no effect of an attention load on URC of single-foreground-object scenes, to URC of scenes with multiple foreground objects. However, Experiments 1 and 2 did not make clear whether or not, using our particularly challenging attention task, URC performance would also be impaired for scenes with only one primary foreground object. Experiment 3 addressed this issue.

There were three types of condition in this new study. To provide a baseline against which to compare our other conditions, Condition A presented, in addition to the stimuli for the attention task used in Experiment 2B, single objects with only their immediate texture and color backgrounds against which they were presented in Experiments 1 and 2. The objects, although not presented as part of a scene, also appeared in the same positions on the screen as they had in Experiment 2A. In Condition B, we presented scenes with a complete scene background but only one primary foreground object. In this respect, these scenes were comparable to those employed by previous URC studies, including those of Li et al. (2002). Finally, in Condition C, we included a condition in which a scene background was presented with four primary foreground objects (as in Experiments 1 and 2).

We predicted that URC performance under dual-task conditions for scenes with single foreground objects would be comparable to that for objects presented in isolation. In contrast, we expected to find that dual-task performance in Condition C where the scenes comprised multiple foreground objects would be substantially impoverished relative to these other two conditions. As an additional measure to ensure that such predicted differences did not simply reflect the different physical properties of the different displays in Conditions A, B, and C, a second group of participants also performed the URC tasks in Conditions A, B, and C (run order counter-balanced as for the first group). The second group of participants saw the same physical displays as the first group, but performed the URC task alone.

### Experiment 3

#### Method

**Participants**

Twelve new participants (two male, ten female) performed three blocks (one block for each of the Conditions A, B, and C) of 40 trials under dual-task conditions. An additional group of six participants (two male, four female) performed the URC tasks of Conditions A, B, and C under single-task conditions.

**Stimuli, apparatus, and procedure**

All aspects of stimuli, apparatus, and procedure were identical to those in the dual-task conditions of Experiment 2A, except that the background scenes took three different forms. An example of each type of scene is illustrated in Figure 5. In Condition A (see Figure 5A), “scenes” consisted of a single object and its immediate background presented in the same locations as they had in the fully “naturalistic” scenes of Experiments 1 and 2. Images were created by isolating all 48 individual objects from the original natural scenes, ensuring each object’s background remained consistent across conditions. In Condition B (see Figure 5B), scenes contained entire backgrounds (as they had in Experiments 1 and 2) but only a single object; in the same position, it would occur in a four-object scene. These scenes are the most akin to those used by previous studies, including Li et al. (2002). Finally, Condition C (see Figure 5C) employed scenes with four major foreground objects as in Experiments 1 and 2.

In creating four-object (rather than, e.g., eight-object) scenes, we had hoped to exhaust *but not to exceed* the...
well-documented (though not uncontroversial) four-object limit on attention and VSTM. Accordingly, if we added an additional attention load (in an attention or VSTM task), then this would exceed the (approximate) four-object limit. Similarly, in our attention and VSTM tasks, we also employed four objects. Accordingly, if a participant attended these objects, their attention and VSTM capacity should have been exhausted, and they should not have capacity to process any further items in the URC task. We thus expected a minor impoverishment in URC in the dual-task compared to the single-task conditions even for one-object scenes (as the attended objects across the two tasks now summed to five objects). This prediction was already in contradiction of the conclusions of Li et al. (2002, 2005) that URC can arise in the virtual absence of attention. Moreover, we predicted a much greater impoverishment under dual-task conditions relative to single-task conditions in the four-object scenes as the number of items to be attended across the two tasks would now be eight objects, far exceeding the putative four-object limit.

Results

Figure 6 graphs percentage accuracy for dual-task conditions (gray bars) and single-task conditions (black bars) for each of the types of scene employed (Conditions A, B, and C indicated by left hand, center, and right hand pairs of bars, respectively). Note under dual-task conditions, performance in Condition C (four-object displays) was substantially lower than that for Condition B (single-objects in full-scene backgrounds), which yielded a similar level of accuracy to Condition A (presentation of a single object and its immediate background). Additionally, note that for single-task conditions, this pattern did not hold.

A mixed two-way ANOVA with factors of task condition (dual-task versus single-task) and scene condition (single object and immediate background, single object in full-scene background, and four-object scenes) yielded a significant main effect of both task condition \((F(2,15) = 5.955, p < 0.05)\), a main effect of scene condition \((F(1,16) = 13.787, p < 0.01)\), and most importantly for our purposes, a significant interaction between these two factors \((F(2,18) = 7.555, p < 0.01)\).

To reveal the source of this interaction, we repeated the above two-way ANOVA three times, excluding Condition A, Condition B, or Condition C and examining the interaction term in each case. For each pair of conditions, a significant interaction between the two factors would confirm that the effect of dual-task versus single-task conditions was greater in one type of scene condition than in the other. The first of these ANOVAs, comparing performance in single-object-plus-background scenes (Condition B) versus single objects in immediate background (Condition A) yielded no significant interaction term \((F(1,16) = 0.588, p = 0.454)\), indicating that that dual-task conditions had affected performance equivalently in the two cases. The deterioration in performance for dual-task relative to single-task conditions in the four-object scenes as the number of items to be attended across the two tasks would now be eight objects, far exceeding the putative four-object limit.

Figure 6. Indicates % accuracy in single-task (dark-shaded bars) and dual-task (light-shaded bars) conditions for the three background conditions in Experiment 3. Left-most pair of data points indicates performance in Condition A (single object and local background), the middle pair of data points indicates performance in Condition B (single object in full-scene background) and the right-most pair of data points performance in Condition C (four-object scenes). Error bars signify 1 SEM.
conditions had not been significantly greater for single-object-plus-background scenes (Condition B) than for a single object presented. In contrast, when the four-object scenes (Condition C) were compared (using identical analyses) to either of the single-object conditions, this yielded highly significant interaction terms for both cases (Condition A vs. Condition C: $F(1,16) = 11.172, p < 0.01$, Condition B vs. Condition C: $F(1,16) = 13.237, p < 0.01$).

Pairwise $t$-tests were then conducted to assess whether these interaction terms reflected performance differences between the conditions under dual-task or single-task conditions. Under dual-task conditions, accuracy in Condition C (four-object scenes) was significantly lower than either Condition A (single-object in isolation; $t(11) = 5.452, p < 0.01$) or Condition B (single-object-plus-background; $t(11) = 4.426, p < 0.01$), whereas performance in Conditions A and B did not differ ($t(11) = 0.645, n.s.$). In contrast, no differences between any two conditions approached significance under single-task conditions, with any numerical trend in the opposite direction to that observed under dual-task conditions (all $t$’s $< 1$, n.s.).

**Discussion**

As we had predicted, scenes comprising a single object against a complete scene background (Condition B) could be categorized under dual or single task conditions as accurately as an object presented in isolation (Condition A). This is consistent with Li and colleagues’ (2002, 2005) earlier findings, suggesting that such displays can be categorized effectively in the “near absence” of attention. However, we found that under dual-task conditions, URC performance was impoverished for scenes comprising four primary foreground objects relative to either of the other conditions. This pattern of results did not arise under single-task conditions indicating that the differences in performance found for dual task conditions did not simply reflect physical differences in the displays. This finding is consistent with the results of Experiments 1 and 2, suggesting that categorization of scenes with four primary foreground objects does require attention resources.

One remaining discrepancy between our findings from those of Li et al. (2002, 2005) is that our attention-demanding tasks reduced URC performance even in single-object-plus-background scenes. However, this effect was weak, only reaching significance on a one-tailed $t$-test. Moreover, when we compared the effects of dual-task versus single-task conditions for our single-object-plus-background scenes (Condition B) to single objects presented in isolation (Condition A) using a two-way ANOVA there was no significant interaction. That is, the impoverishment of URC performance under dual-task conditions was not significantly greater for single-object-plus-background scenes that for a single object presented in isolation. To a certain extent therefore, our findings do support the Li et al. conclusions; some processes involved in URC of one-object-plus-background scenes (namely those involved in negating the effects of the scene background, possibly by guiding attention to the primary foreground object) do seem to arise even when participants are trying to attend to a central task. However, in our studies, URC as a whole does seem to be dependent on the availability of attention resources, even for one-object-plus-background displays.

How might we explain the partial discrepancy between our findings and those of Li and colleagues? One possibility is that because our objects were repeated within the experiment coding of scenes in later trials might have suffered interference due to common object identities and key locations they shared with earlier trials in a session. It is conceivable, in such a case, that this interference may have made URC in our tasks harder and therefore more attention-demanding than it might otherwise have been. However, we also believe that such destructive interference is unlikely to have played a major role in our findings, given the small number of times that each object was presented, particularly in Experiment 3 (1.67 times per condition). Additionally, we note that the surprising element of URC performance is that it can arise even though each scene has never been seen before. Once participants are familiar with shapes or contexts, they (in general) are searched more efficiently (Flowers & Lohr, 1985), guide attention more efficiently (e.g., Chua & Chun, 2003), and are categorized more efficiently (Valentine, 1991) than novel objects. Any repetition should have made our task easier and less attention demanding, not more so.

Instead, we are sure that the partial discrepancy relates to the different attention-demanding tasks in the two studies. Our attention-demanding vowel-detection task averaged around 70% accuracy across our experiments and was able to elicit reductions in URC performance even for objects presented in isolation. In contrast, the Li et al. (2002) task was titrated to 80% accuracy. This suggests that the Li et al. search task may not have been as attention-demanding as the task we employed in Experiment 3. Indeed, the only empirical evidence that their task could yield measurable deficits in any task was found for briefly presented, abstract stimuli whose edges were immediately masked. We therefore suspect that the Li et al. attention task (a search task) may not have been particularly demanding of attention resources—were they to have used a more attention-demanding task, we are sure that they too would have found a minor deterioration in URC for single-object-plus-background scenes.

**General discussion**

The current findings demonstrate that URC of naturalistic scenes with multiple foreground objects requires attention resources and is substantially impaired when
attention resources are devoted to a vowel-detection or VSTM task. In contrast, and in partial confirmation of previous work, we found that URC of scenes comprising only one major foreground object was attenuated only mildly when attention resources were limited to approximately the same extent as categorization of single objects presented in isolation. Together, these two sets of findings suggest that although simple scenes can be categorized using relatively few attention resources, more complex scenes, typical of those encountered in nature, cannot. This impairment in URC of complex scenes was observed here even when each to-be-categorized scene was presented for 500 ms—18 times as long in the Li et al. (2002) study. Accordingly, it would appear unlikely that URC performance in our studies suffered as the result of brief scene presentations.

Although we make no specific claims here about the capacity of visual attention, our findings in Experiment 3 (and in the other studies) can readily be explained using the following two assumptions. First, it seems plausible that any available attention resources will tend to be drawn to salient foreground objects due to the physical properties of their images on the retina. Second, it has been suggested that attention and VSTM may, for some purposes, select four objects at a time (see e.g., Luck & Vogel, 1997).

The first of these assumptions explains why, in Experiment 3, the scenes with one primary foreground object exhibited equivalent URC performance to objects presented in isolation. If any available attention resources devoted to the scene are reflexively cued to the position of the major foreground object in the single-object-plus-background scenes, then participants will in essence not even process the background. Hence, we should expect to find no differences in performance between these scenes and single objects presented in isolation.

The second assumption explains why under single task conditions, four-object-plus-background scenes can be categorized as efficiently as single-object-plus-background scenes. Provided that (according to the first assumption) attention resources are drawn to the primary foreground object(s) in a scene, neither the four-object nor the one-object scenes will exceed our attention/VSTM capacity of four-objects.

This second assumption also accounts for the much greater subsequent impairment in performance for four-object than one-object scenes and for the presence of a mild impairment in performance for one-object scenes under dual-task conditions. In the case of four-object scenes, the presumed four-object capacity of attention completely inadequate to cope with the eight relevant objects presented (four in the attention task plus four in the scene). For the one-object scenes, in contrast, the presumed four-object capacity of attention resources would have been only slightly exceeded as five task-relevant objects were now presented (four in the attention task plus one in the scene). Accordingly, we would suggest that the detail of our findings is consistent with allocation of limited processing resources to URC that have a maximum capacity of around four objects.

In terms of extant explanations of URC, our findings provide evidence against possible accounts of URC in which high-level processing of animal identities arises efficiently in parallel for many objects in a scene simultaneously. Rather, we would suggest that our results are most supportive of Evans and Treisman’s (2005) conclusions reviewed earlier, which suggest that accumulation of evidence from many low-level visual processors concerning the presence of typical animal features may underpin URC. It would seem likely that such a process might work efficiently for scenes in which there is one primary candidate object. However, the inclusion of multiple candidate objects would be expected to make distinguishing one scene from another on the basis of such features more difficult due to the inclusion of objects with similar spatial frequency spectra, textures, and colors to the target stimuli.

While we favor the Evans and Treisman (2005) account over the assumption of massively parallel, high-level processing in natural scenes, our conclusions here are equally consistent with a third, less palatable explanation. If URC only arises efficiently in natural scenes in which an observer’s attention is drawn to four (or fewer) primary foreground objects due to those objects’ physical image properties, then URC may simply result from this process and detection of a single target (animal or vehicle) from among the few attended objects. Such processing would closely parallel that found within simple abstract arrays of objects (e.g., Duncan, 1980) and would threaten assumptions in the literature that URC of natural scenes is different to that in 2-D abstract arrays of objects. Future research should attempt to distinguish between these possibilities.

One potential objection to our conclusions might be to suggest that the speed with which URC processes can be executed (e.g., Kirchner & Thorpe, 2006) means that they must be able to arise in the absence of attention. However, high-speed processing does not by itself rule out the necessity of attention for performing a task—even the simplest “pop-out” searches require attention (Joseph, Chun, & Nakayama, 1997). Additionally, we would suggest that because those studies use scenes in which the animal is typically the only (or primary) object, such findings are probably more informative about processing of several (four or fewer) attended objects than natural scenes per se. This is not, of course, to imply that such studies do not tell us interesting or novel things about vision. Rather, we suggest that these processes require limited-capacity attention processes and may not be specific to natural scenes.

In summary, our findings challenge the proposal that URC in general can occur in the “near-absence of attention” and suggest that URC, like other high-level visual processes is subject to capacity restrictions. Accordingly, it would seem that URC does not run counter to the prevailing wisdom in the cognitive neuro-
sciences that high-level visual processing is relatively inefficient in cluttered, attention-demanding environments. Contrary to the promises of early work on URC, this phenomenon does not seem to reflect a high-level visual process that can operate independently of attention. Rather, on closer inspection, URC findings are more in keeping with the classical view that only relatively elementary scene properties can be processed efficiently in parallel for many items simultaneously. Braun (2003) claimed that URC findings “upset the visual applecart” in that URC can arise rapidly and independently of attention. If this was the case, then the current findings go some way towards returning the applecart to its previous state.

Acknowledgments

This research was supported by an MRC studentship awarded to S. Walker from 2005 to 2008.

Commercial relationships: none.

Corresponding authors: Sarah Walker or Greg Davis.

Email: sw345@cam.ac.uk or gjd1000@cam.ac.uk.

Address: Department of Experimental Psychology, Downing Street, Cambridge, CB2 3EB, United Kingdom.

References


