ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research xxx (2005) xxx–xxx

Vision Research

www.elsevier.com/locate/visres

Binding is a local problem for natural objects and scenes

Rufin VanRullen [a,*], Lavanya Reddy [b], Li Fei-Fei [c]

[a] *Centre de Recherche Cerveau et Cognition, CNRS-UPS, 133 Rte de Narbonne, 31062 Toulouse Cedex, France*
[b] *CNS Program, Division of Biology, California Institute of Technology, MC 139-74, Pasadena, CA 91125, USA*
[c] *Division of Electrical Engineering, California Institute of Technology, MC 139-74, Pasadena, CA 91125, USA*

Received 13 November 2004; received in revised form 18 May 2005

Abstract

Current theories hold that attention is necessary for binding the features of a visual object into a coherent representation, implying that interference should be observed when two objects must be recognized simultaneously: this is the well-known binding problem. Recent studies have suggested, however, that discriminating isolated natural scenes, objects or faces might be possible in the near absence of attention. It is still unclear what mechanisms underlie this remarkable ability. Here, we investigate whether the binding problem affects natural objects in the same way as other stimuli: is interference observed when two natural objects or scenes must be simultaneously processed? We show that in the presence of competing objects, performance in the near absence of attention depends on the relative distance between stimuli: discrimination is good for stimuli far enough apart, and poor for close enough stimuli. In contrast, seemingly simpler but unfamiliar synthetic objects could not be bound in the near absence of attention, independent of the distance between them. Thus, natural objects are special in that they suffer from the binding problem, but only locally. We surmise that this particular type of local binding for natural objects and scenes could be "hardwired" by dedicated neuronal populations.

Keywords: Attention; Binding; Natural scenes; Object recognition; Visual research; Dual-task

1. Introduction

A substantial area of visual neuroscience research concerns the features of visual objects that can be detected "preattentively". It is thought that these features constitute the building blocks that can be bound together, under the effect of attention, to compose our mental representations of objects: the "Feature-Integration Theory" (Treisman & Gelade, 1980). According to this theory, the color, shape, motion and other basic properties of an object are only linked after directed attention permits the creation of a specific "object file". Parietal cortex is likely to play a key role in directing attention for feature binding (Ashbridge, Cowey, & Wade, 1999; Friedman-Hill, Robertson, & Treisman, 1995; Shafritz, Gore, & Marois, 2002). Central to the feature integration view is the idea that object representations do not exist outside the focus of attention. This is supported by numerous visual search experiments demonstrating that, although simple features such as color, orientation, motion direction and so on, are indeed "preattentive", higher-level properties of objects such as their identity or category seem to require focused attention (Wolfe, 1998). The critical variable in the visual search paradigm is the dependence of performance or reaction time on the number of stimuli simultaneously presented (set size): if the target feature can be detected independent of set size, then it is said to be preattentive; if not, then it is assumed that each item in the display had to be explored with the attentional focus.

A critical, but often overlooked issue is that the crowded displays used in visual search might confound the attentional requirements of object recognition per

[*] Corresponding author.
E-mail address: rufin@klab.caltech.edu (R. VanRullen).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.05.012

se with those induced by the competition among neighboring stimuli (VanRullen, Reddy, & Koch, 2004). In other words, serial search might arise from the need to focus attention on each item, not in order to bind ‘‘preattentive’’ features, but in order to resolve this spatial competition (Desimone & Duncan, 1995; Reynolds & Desimone, 1999). In this case, the binding process itself might very well occur ‘‘preattentively’’. In fact, it was found recently that attentional requirements observed using isolated objects (such as in dual-task paradigms) do not match those obtained in visual search (VanRullen et al., 2004). For example, in the near absence of attention it is possible to determine the presence of an animal in an isolated natural scene (Li, VanRullen, Koch, & Perona, 2002), or to determine the gender of an isolated face (Reddy, Wilken, & Koch, 2004). The same tasks are not performed efficiently in visual search (VanRullen et al., 2004), presumably due to competition within the displays (Reddy, VanRullen, & Koch, 2005). These recent findings suggest that object identity or category might in fact be represented ‘‘preattentively’’. Does this imply that, contrary to the influential Feature Integration Theory, the binding of object features can sometimes occur outside the focus of attention? Not necessarily. It could also be that this ability relies on a ‘‘default’’ binding strategy, whereby the mere presence of the object features at any location in the visual field, possibly disjoined, would be sufficient for recognition. For ‘‘true’’ binding to occur, the object features must be detected simultaneously within the same area of visual space.

To illustrate this, imagine a task involving the detection of any *red* object moving *rightwards*. Leftward-moving red objects, or rightward moving objects of a different color, would constitute distractors for this task. As long as only isolated stimuli are involved, this task can be solved by monitoring ‘‘red’’ feature detectors and ‘‘rightward’’ feature detectors: when both features are present simultaneously, the target can be safely detected. This is essentially a ‘‘default’’ binding strategy, which does not require object features to be detected at the *same* location. If a white object moving rightwards is shown simultaneously with a stationary red object, however, this strategy will lead to a false detection (the so-called ‘‘illusory conjunction’’). This is the well-known ‘‘binding problem’’, proposed to occur whenever attention cannot simultaneously handle all objects in the field (Treisman & Schmidt, 1982).

An indirect way to test binding mechanisms for natural objects and scenes could thus be to use a similar form of ‘‘illusory conjunction’’ paradigm. Adding a task-irrelevant object in the visual field on each trial (in addition to a target, task-relevant object) can be a good way to ensure that certain object features are always present. A ‘‘default’’ binding strategy would predict that the irrelevant object would interfere with the target on each trial, even though it is not relevant to the task. On the other hand, ‘‘true’’ object binding would mean that the features of the target (task-relevant) object are bound in a spatially specific manner, and do not suffer interference (i.e., ‘‘illusory conjunction’’) from the features of the irrelevant object—if it is placed far enough away. To sum up this line of reasoning, whereas ‘‘default’’ binding predicts interference between object features in the absence of attention, ‘‘true’’ object binding would predict no interference when objects are spaced sufficiently. This latter result is precisely what we report for natural objects and scenes in two series of experiments.

Our aim was to determine if, and under what conditions, natural objects and scenes are affected by the binding problem. We define this problem as a decrease in performance (i.e., ‘‘interference’’ or ‘‘illusory conjunction’’) occurring specifically when two stimuli must be simultaneously processed. In the first series of experiments, the dual-task paradigm is used with natural target categories (i.e., animal vs. non-animal scenes, or upright vs. inverted faces), allowing us to control for the allocation of focal attention. On some trials a small, ‘‘distracting’’ picture (an animal scene in the former case, an upright face in the latter) is added to the display, either close (same quadrant) or far (opposite quadrant) from the target stimulus. The subjects are instructed to ignore this small picture, which they manage quite well when attention is available (‘‘single-task’’ condition). When attention is occupied elsewhere, however (i.e., in the ‘‘dual-task’’ condition), strong interference is observed if the ‘‘distracting’’ picture is close to the target, but little or no interference when it is far. Thus, in the near-absence of attention, our natural stimuli are only affected by a *local* binding problem.

In the second series of experiments, a comparison task is used instead of a dual-task. Two stimuli are presented simultaneously (animal or non-animal scenes, upright or inverted faces), either close or far from one another, and must be compared. The subjects' recognition performance on each stimulus in isolation is known (‘‘reference’’ performance), and thus the optimal performance for the comparison task can be predicted: it corresponds to the performance that would be obtained if both stimuli were recognized at the ‘‘reference’’ level, i.e., if there was no interference. Here again, near-optimal performance is observed for distant stimuli, while interference appears for close stimuli. Importantly, we verified that other, synthetic stimuli (randomly rotated letters, bisected 2-color disks), whether close or far from each other, systematically undergo significant interference under the same conditions. This is again in favor of a purely local binding problem affecting the processing of natural objects and scenes—in striking contrast to artificial geometric shapes, for which binding appears to depend on a more global resource.

2. Methods

2.1. General procedure

Subjects were seated in a dimly lit room at approximately 120 cm from a computer monitor (refresh rate 75 Hz) piloted from a PC computer. The experiments were programmed using the Presentation software. Display timing accuracy was controlled a posteriori for each trial. All subjects provided informed consent before participation. Over the course of this study, seven new "naïve" subjects participated in experiments 1 and 2. The rest consisted in seven previously trained subjects who had participated in some of our previously published experiments. It is thus important to note that the main conclusions of our manuscript still hold when considering only the set of "naïve" subjects.

2.2. Experiment 1

2.2.1. Subjects

Ten subjects (two authors, four undergraduate and graduate students from the California Institute of Technology in Pasadena, USA, plus four naïve undergraduate and graduate students from the Centre de Recherche Cerveau et Cognition in Toulouse, France) previously trained in dual-task performed the animal vs. non-animal scene categorization experiment. Another four subjects (one author, plus three naïve undergraduate and graduate students from the Centre de Recherche Cerveau et Cognition) previously trained in dual-task performed the upright vs. inverted face discrimination experiment.

2.2.2. Central task: 5-letter discrimination

Each trial started with the appearance of 5 letters (randomly rotated Ls and Ts) randomly occupying 5 out of 9 possible positions at the center of the screen. Each letter measured less than 0.5 deg of visual angle, and their maximum eccentricity was 1 deg. On half of the trials, the letters were all the same (5 Ls or 5 Ts). On the remaining half, one letter differed from the other 4. All letters were masked by the letter F using the same size and the appropriate rotation angle. The stimulus onset asynchrony (SOA) was determined individually for each subject to avoid saturation (i.e., keeping performance of this task around 75%). Central letter SOAs ranged from 186 to 240 ms for different subjects. The same SOAs were used in the single-task and dual-task experimental blocks.

2.2.3. Peripheral task: Animal vs. non-animal scene categorization

In this version of the experiment, the peripheral stimuli were colored natural scenes and subjects had to determine whether the scene contained an animal or not. These stimuli were obtained from a large commercial database, and were similar to those used in previous studies (Li et al., 2002; Thorpe, Fize, & Marlot, 1996; VanRullen et al., 2004). Stimuli were masked using a combination of noise filtered at various spatial frequencies, on which a colored texture was superimposed (see Figs. 1 and 3 for examples). The stimulus-onset asynchrony (SOA) was determined separately for each subject to yield a performance of 70–85% correct. In all cases, the peripheral stimulus was masked before the end of the central letters SOA. Peripheral SOAs for natural scenes ranged from 80 to 160 ms. The same SOAs were used in the single-task and dual-task experimental blocks.

2.2.4. Peripheral task: Upright vs. inverted face discrimination

In this version of the experiment, the peripheral stimuli were upright or inverted faces and subjects had to discriminate the orientation (upright vs. inverted). There were 2 grayscale upright face stimuli (one male and one female), and the same faces were used as inverted face stimuli. Stimuli were masked using a grayscale mixture of geometrical shapes (see Figs. 4 and 5 for an example). The stimulus-onset asynchrony (SOA) was determined separately for each subject to yield a performance of 70–85% correct. In all cases, the peripheral stimulus was masked before the end of the central letters SOA. Peripheral SOAs ranged from 53 to 133 ms for face stimuli. The same SOAs were used in the single-task and dual-task experimental blocks.

2.2.5. Interference conditions

Peripheral stimuli in the dual-task paradigm were presented under 3 equiprobable interference conditions (Fig. 1). In the "no interference" condition, the peripheral stimulus was shown at a randomly determined position on a virtual rectangle at approximately 5.5 deg of eccentricity. This peripheral stimulus, measuring 4×3 deg of visual angle, was presented alone, without an additional distracting stimulus (the 5 central letters, however, remained on the screen throughout the peripheral stimulus presentation, and until after the appearance of the peripheral mask; see Fig. 1). In the "close interference" condition everything happened in exactly the same way, except that a smaller (about 3×2 deg of visual angle) stimulus was simultaneously presented, halfway between the central letters and the peripheral stimulus (i.e., on average at 2.5 deg eccentricity and at a distance of 3 deg from the peripheral stimulus, center-to-center). Thus the additional stimulus and the peripheral task-relevant stimulus always belonged to the same quadrant (except in the occasional case when they both lay on the horizontal or vertical meridian). In the "far interference" condition, the additional stimulus was placed diametrically opposite to the location it would have occupied in the "close interference" condi-

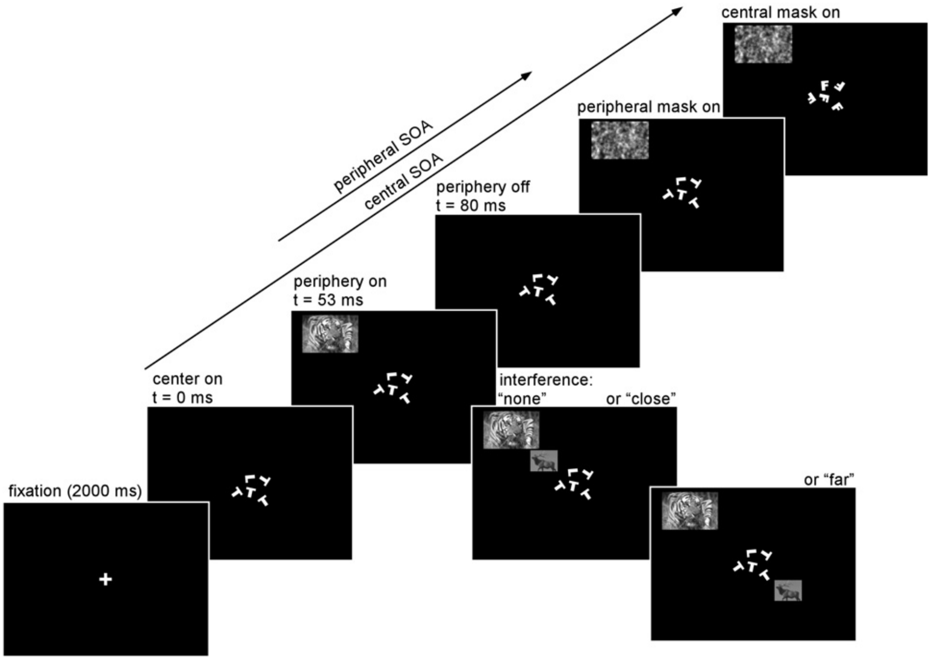

Fig. 1. Dual-task paradigm and the three interference conditions. We used a modified version of the dual-task paradigm, in which a small, ''interfering'' task-irrelevant stimulus was added randomly on 2/3 of the trials. When present, this stimulus was placed (with equal probability) either close to the larger, task-relevant peripheral stimulus (at a distance of about 3 deg), or far from it (at about 8 deg). For each subject, this interfering stimulus was kept identical throughout the entire experiment, while the task-relevant stimulus, as well as its position, changed randomly on each trial. Similar trials were shown in blocks of 96 trials with varying task instructions. In the single central task, the subjects determined whether the 5 randomly rotated letters (Ls and Ts) were all identical or not, and ignored peripheral stimuli. In the single peripheral task condition, subjects ignored the central letters and had to discriminate the large peripheral stimulus. For one group of subjects, the stimulus was a colored natural scene and subjects decided whether an animal was present or not (illustrated here). For another group, the stimulus was a grayscale face and the subjects determined its orientation (upright vs. inverted). They were explicitly instructed to ignore the occasional smaller, interfering stimulus. Finally, in the dual-task condition subjects were required to perform both tasks simultaneously, maintaining attention on the central letters.

tion. In other words, the additional stimulus and the peripheral task-relevant stimulus always belonged to opposite quadrants (on average 8 deg apart, center-to-center).

In both peripheral tasks (natural scene categorization or face orientation discrimination), subjects were instructed to categorize the larger, peripheral stimulus and ignore the occasional smaller additional stimulus. For any given subject, the additional stimulus was kept identical throughout the entire experiment. Thus, the observers were given every chance to optimize whatever mechanisms might allow them to successfully ignore this irrelevant stimulus. In the ''animal vs. non-animal scene'' experiment, the additional stimulus was an ''animal'' scene, different for each subject. In the ''upright vs. inverted face'' experiment, the additional stimulus was an upright face. To maximize interference, the additional stimulus was not masked, and remained visible until the subject's response.

2.2.6. Instructions

Each subject performed at least 3 one-hour sessions for these experiments. Each session comprised several randomly interleaved blocks of 96 trials of the single central task, the single peripheral task and the dual-task conditions. All trials contained both a central and a peripheral stimulus, and the specific instructions determined which was relevant for the current block. In the *single central task* condition, subjects were instructed to focus attention on the central 5 letters (randomly rotated Ls and Ts) and determine whether they were all the same or whether one of them differed from the other four. This task has been repeatedly demonstrated to efficiently engage focal attention (Braun & Julesz, 1998; Li et al., 2002; Reddy et al., 2004; VanRullen et al., 2004). In this condition, subjects were free to ignore the peripheral stimuli. In the *single peripheral task* condition, subjects were instructed to ignore the central letters, and discriminate the larger, more peripheral stimulus (animal vs. non-animal, or upright vs. inverted face). They were also warned that an occasional, smaller distracting stimulus might appear, closer to fixation, and were instructed to disregard it. The performance of the subjects in these two single-task conditions (further separated according to the 3 ''interference'' conditions: ''no'', ''close'' or ''far'' interference) served as reference points to estimate the dual-task performance. In the *dual-task* condition, subjects were instructed to keep their attention focused on the central 5 letters, and perform this task with maximal accuracy. At the same time, they were required to provide a response on the peripheral (task-relevant) stimulus.

2.2.7. Performance normalization

As in other dual-task studies, we estimate dual-task performance not in terms of absolute levels, but with respect to the corresponding single-task performance. For each subject, this normalization transforms the average dual-task performance d into a normalized performance d_{norm} for which 50% represents chance and 100% represents the corresponding single-task performance s:

$$d_{\text{norm}} = 0.5 + 0.5 \ * (d - 0.5)/(s - 0.5).$$

The same normalization is applied to both the central and peripheral tasks performances in the dual-task condition, leading to the data shown in Fig. 2. Importantly, the single-task performance levels used as references for this transformation were calculated separately for each of the 3 interference conditions (“no”, “far” and “close” interference).

2.3. Experiment 2

2.3.1. Subjects

Five subjects (including one author and four naïve subjects) participated in the main part of experiment 2. Four subjects (one author, one subject from the main part of experiment 2, and two additional subjects) participated in the control experiment.

2.3.2. Discrimination tasks

We used four discrimination tasks in this paradigm: a natural scene categorization task, an upright vs. inverted face discrimination task, a L/T letter discrimination task, and a bisected disk discrimination task. For the natural scene categorization and the upright vs. inverted face discrimination tasks, stimuli and masks were similar to those used in experiment 1. In the letter discrimination task, the stimuli to be discriminated were single, randomly rotated Ls and Ts, masked by a letter F rotated appropriately. In the bisected disks task, stimuli were green- red or red-green vertically bisected disks, masked by disks having alternating red and green quadrants. One subject’s data from the letter and bisected disk tasks were discarded a posteriori, after she reported using apparent motion between stimulus and mask to perform the tasks.

2.3.3. Single-stimulus discrimination

The stimulus, measuring about 3 deg of visual angle, was presented randomly at one of 8 possible locations (of equal eccentricity at 5 deg). In each task the SOA was varied randomly between 27 ms and 213 ms. This allowed us to trace, for each subject and each of the four discrimination tasks, a psychometric curve (based on the normal probability density function). Using this fit we could precisely determine the SOA necessary for 85% correct performance (Fig. 3B).

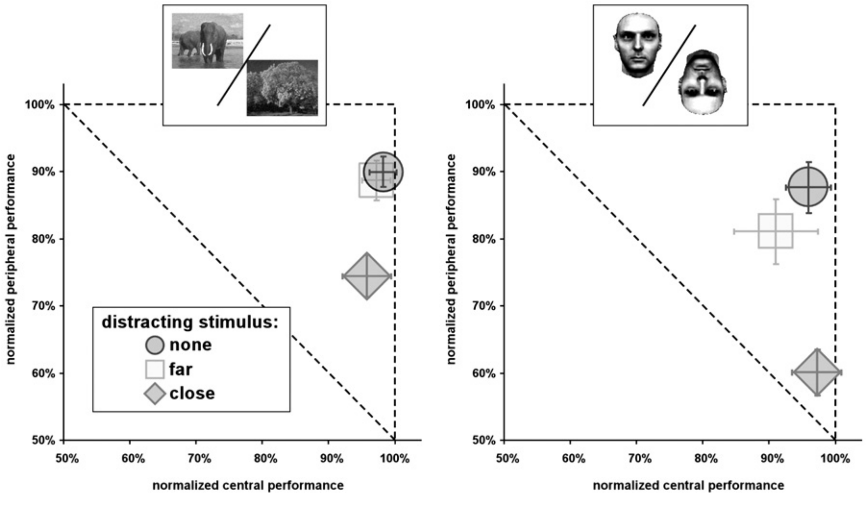

Fig. 2. Normalized dual-task performance for the animal vs. non-animal scene categorization (left) and the upright vs. inverted face discrimination tasks (right). On both axes, the performance is plotted relative to the corresponding performance obtained in the single-task condition (for trials of the identical “interference” condition). Each point corresponds to the average dual-task performance over the entire group of subjects (10 observers for the data in the left panel, 4 for the data shown in the right panel). Error bars represent standard error of the mean. There is a significant main effect of interference condition ($p < .001$ for the natural scene task, $p < .005$ for the face task). In the “no interference” condition (circles), natural scene categorization or face orientation discrimination are performed fairly well when attention is unavailable (about 90% of the corresponding single-task performance for the natural scenes, and more than 85% for the faces). In the “far interference” condition (squares), the additional stimulus impairs performance only minimally (post-hoc test, $p > .05$). In the “close interference” condition (diamonds), significant interference is observed for both groups of subjects ($p < .05$). This interference is taken as the signature of a binding problem, which in the present case is only observed locally.

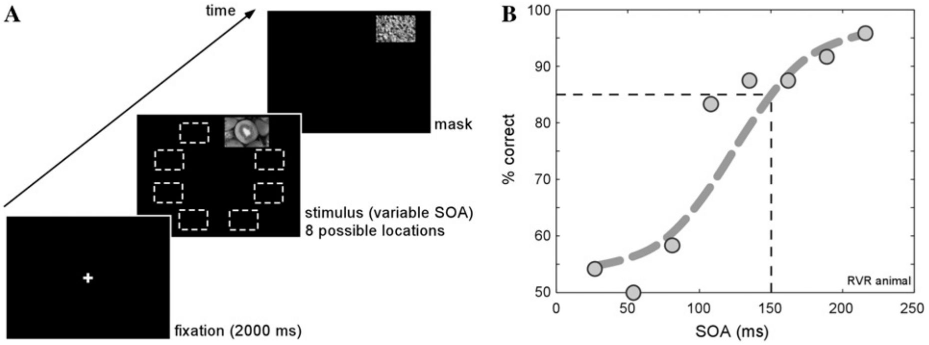

Fig. 3. Obtaining psychometric curves for isolated stimuli. (A) One single stimulus (measuring less than 3 deg of visual angle) was presented randomly at one of 8 possible locations (same eccentricity of about 5 deg) and had to be categorized. 5 subjects performed this experiment using 4 different tasks (192 trials were run in each task): animal vs. non-animal natural scene categorization (illustrated here), upright vs. inverted face discrimination, letter discrimination (randomly rotated L or T) and 2-color vertically bisected disk discrimination (red-green vs. green-red). Stimuli were masked and the SOA was varied systematically between trials. We recorded discrimination performance as a function of SOA. Figure not to scale. (B) For each of the subjects and tasks, we fitted the data using a psychometric function (shown here for one subject in the natural scene categorization task), and used the fit to determine the exact SOA at which a performance of 85% correct could be expected. This SOA was then used in the subsequent comparison task.

2.3.4. Comparison task

In the second experimental phase, stimuli were presented in pairs and had to be compared (“same” or “different” category). They were shown in one of two possible spatial configurations. In the “far” condition, the two stimuli were presented in opposite quadrants, 8 deg apart, at the same eccentricity; in the “close” condition, they were shown in the same quadrant, 3 deg apart, again at the same eccentricity. The quadrant(s) was (were) chosen randomly for each trial. Overall, the stimuli in this second phase had the same distribution of spatial positions as in the previous phase. Because the SOA allowed for 85% correct performance in that first phase, it is easy to calculate what performance level is expected in the second phase if the two stimuli can be recognized without interference. In the best case, assuming that each isolated stimulus, irrespective of its position, could be recognized at 85% correct, one should obtain a comparison performance of $0.85^2 + (1-0.85)^2 = 0.7225 + 0.0225 = 0.745$. (Correct performance on a given trial will be obtained either if both stimuli are correctly categorized, or if both are wrongly categorized.) If one assumes that some stimulus positions lead on average to a better performance than others, one can estimate that in the worst case (with some positions leading to 100% correct performance and others to only 70%, so the total average performance remains 85%) one should obtain a comparison performance of $1 * 0.70 + (1-1)*(1-0.70) = 0.70$. Overall, “optimal performance” in the comparison task should thus lie between 70% and 75% correct.

2.3.5. Control experiment

We replicated the previous comparison experiment (including the first phase used to determine psychometric curves) with stimulus pairs constrained to one visual hemifield (randomly determined for each trial). Here, the stimulus eccentricity was increased slightly compared to the previous experiment, so that inter-stimulus distances remained comparable (at more than 8 deg and less than 3 deg for the “far” and “close” conditions, respectively).

3. Results

3.1. Experiment 1: Competition in dual-task

In a dual-task paradigm, even when attention is taken away by a central letter task (discriminating 5 randomly-rotated Ts and Ls), it is still possible to categorize natural scenes (animal vs. non-animal, vehicle vs. non-vehicle; Li et al., 2002) or faces (male vs. female; Reddy et al., 2004) in the periphery. What would happen if competition were introduced in this paradigm, i.e., if a task-irrelevant object was introduced in addition to the scene or face to be categorized? If this type of “high-level” discrimination without attention relies on a “default binding” strategy, then strong interference should be observed. If, on the other hand, this ability relies on true (spatially specific) object binding, the additional object should not interfere with the task-relevant object–at least as long as it does not lie within the spatial range of this binding process.

Here peripheral stimuli in the dual-task paradigm were presented under 3 equiprobable interference conditions (Fig. 1). In the “no interference” condition, the peripheral stimulus was presented alone, without an additional distracting stimulus. In the “close interference” condition a smaller stimulus was simultaneously presented, halfway between the central letters and the peripheral stimulus (on average at 3 deg from the peripheral stimulus). Thus, the additional stimulus and the peripheral task-relevant stimulus always belonged to the same quadrant. In the "far interference" condition, the additional stimulus was placed diametrically opposite to the location it would have occupied in the "close interference" condition. In other words, the additional stimulus and the peripheral task-relevant stimulus always belonged to opposite quadrants (on average 8 deg apart, center-to-center).

There were two different versions of this experiment. In one version, the peripheral stimuli were colored natural scenes and subjects had to determine whether the scene contained an animal or not. In the other version, the peripheral stimuli were upright or inverted faces and subjects had to discriminate the orientation (upright vs. inverted). In both tasks (natural scene categorization or face orientation discrimination), subjects were instructed to categorize the larger, peripheral stimulus and ignore the occasional smaller additional stimulus.

All trials in this experiment comprised both a central (5 letters) and a peripheral stimulus as shown in Fig. 1. These trials were shown in separate blocks with varying instructions: subjects were either asked to perform the central task (letter discrimination) alone ("single central task"), the peripheral task (animal vs. non-animal, or upright vs. inverted faces) alone ("single peripheral task") or both tasks simultaneously ("dual-task"). Performance in the two single-task conditions served as reference points to estimate the dual-task performance.

For both the "animal vs. non-animal" and the "upright vs. inverted face" tasks, subjects performed fairly well when only a single peripheral stimulus was presented, even when attention was tied at the center of the screen (Fig. 2). The peripheral performance in the dual-task condition (i.e., in the near absence of attention) lay on average between 85% and 95% of the performance level obtained in the single-task condition (i.e., with attention available). This lack of strong attentional requirement for isolated natural stimuli is to be expected on the basis of our previous results (Li et al., 2002; Reddy et al., 2004). Crucially though, when the peripheral stimulus was presented simultaneously with an additional distracting stimulus, dual-task performance depended on the distance between the two stimuli (one-way ANOVA, effect of the interference condition: "close", "far" or "no" interference; $F(2,27) = 9.45$, $p < .001$ for the animal vs. non-animal task; $F(2,9) = 12.71$, $p < .005$ for the upright vs. inverted face task; note that the degrees of freedom involved in these two tests differ, due to different numbers of subjects): in the "far interference" condition, performance was indistinguishable (Tukey–Kramer post-hoc test, $p > .05$) from that obtained with isolated stimuli; in the "close interference" condition, performance suffered a strong decrease ($p < .05$). Note that the three "normalized" dual-task performances reported here were estimated relative to the single-task performance of the corresponding interference condition ("no", "close" or "far" interference). In other words, these results truly reflect the attentional requirements of the various conditions, and not merely the difficulty of the task itself. This means that crowding or local interference, which would make processing of the peripheral stimulus altogether more difficult in the "close interference" condition—whether in the single- or the dual-task conditions—could not be called upon to explain these results.

To summarize, the distracting stimulus was only found to affect performance–i.e., to produce a binding problem–in the "close interference" condition, when the task-relevant and task-irrelevant natural stimuli belonged to the same quadrant (less than 3 deg apart). As we predicted, this implies that this discrimination relies on a type of object binding that is spatially specific.

Do these results depend on our choice of "distracting" stimulus? Would we obtain opposite results if the additional stimulus belonged to the opposite category (non-animal scene, or inverted face), or an altogether different category (for example, a single bright flash)? Clearly, we cannot answer this question based on the present data, but this is in fact not relevant to the present argument. We chose these particular "distracting" stimuli because we expected that they had the potential to maximize interference–and indeed significant interference was observed. If interference was observed similarly with other types of distracting stimulus, it could simply mean that they are also able to trigger competition with the relevant neural mechanisms. If not, then it could simply mean that they do not compete significantly. What matters for the sake of our argument is that when the neural representations of the distracting stimulus and of the peripheral stimulus do compete, then this competition is observed only locally.

3.2. Experiment 2: Comparison task

An objection often raised against our findings of limited attentional requirements for natural object categorization tasks is that our dual-task paradigm (as in e.g., (Braun & Julesz, 1998; Braun & Sagi, 1990)) uses a central, attentionally demanding task (5-letter discrimination) that has little in common with the peripheral natural discrimination tasks. This might even explain the fact that peripheral synthetic stimulus discrimination tasks (bisected 2-color disks, rotated L vs. rotated T), which resemble the central task, suffer more attentional deficits in this paradigm. This type of objection, however, cannot account for the results of Rousselet, Fabre-Thorpe, and Thorpe (2002), who found that processing two natural scenes at once, one in each hemifield, could be done without interference, as easily as with one single scene (Fei-Fei, VanRullen, Koch, & Perona, 2005). Here, we use a variant of that paradigm, where the respective categories of two simultaneously presented stimuli must be compared. The distance between the two stimuli is varied on each trial. The general idea is that if the features of both stimuli can be independently bound, without interference and thus presumably without attentional requirements, then performance on this comparison task should be close to optimal. On the other hand, if this type of binding is attention-dependent, then only one of the two stimuli could be recognized on each trial, and comparison performance should be fairly low. In essence, this experiment is thus comparable to a visual search experiment: we investigate the effects of increasing set size from one to two simultaneously presented stimuli. In our case however, competition is carefully controlled by always displaying only two stimuli, at various distances; additionally, the task design requires subjects to efficiently process both stimuli: recognizing only one of them does not increase the probability of a correct response. We applied this paradigm to four different tasks: the ‘‘animal vs. non-animal’’ natural scene categorization task, the ‘‘upright vs. inverted’’ face discrimination task (as in Experiment 1), and two tasks that have been repeatedly found to suffer high attentional costs in dual-task paradigms (discrimination of single randomly rotated Ls or Ts; discrimination of bisected two-color disks).

This experiment was performed in two separate phases. In the first phase, discrimination performance was measured on isolated stimuli (Fig. 3). The stimulus was presented randomly at one of 8 possible locations (of equal eccentricity at 5 deg). The stimulus onset asynchrony (SOA) was varied randomly between 27 and 213 ms. This allowed us to trace, for each subject and each of the four discrimination tasks, a psychometric curve, and precisely determine the SOA necessary for 85% correct performance (Fig. 3B). The essential ruse in this experiment was to use exactly this SOA during the second phase. The resulting SOAs (average ± standard deviation across subjects) were 173.5 ms (±13.5 ms) for the animal vs. non-animal scene categorization task; 117.5 ms (±25.5 ms) for the upright vs. inverted face task; 120 ms (±56.5 ms) for the rotated L vs. T discrimination task; and 157 ms (±35 ms) for the bisected disk discrimination task.

In the second experimental phase, stimuli were presented in pairs and had to be compared (‘‘same’’ or ‘‘different’’ category). They were shown in one of two possible spatial configurations (Fig. 4A). In the ‘‘far’’ condition, the two stimuli were presented in opposite quadrants, 8 deg apart, at the same eccentricity; in the ‘‘close’’ condition, they were shown in the same quadrant, 3 deg apart, again at the same eccentricity. The quadrant(s) was (were) chosen randomly for each trial. Overall, the stimuli in this second phase had the same distribution of spatial positions as in the previous phase. Because the SOA allowed for 85% correct performance in that first phase, we can calculate that if the two stimuli can be recognized without interference in the second phase, ‘‘optimal comparison performance’’ should lie between 70% and 75% correct (see Section 2).

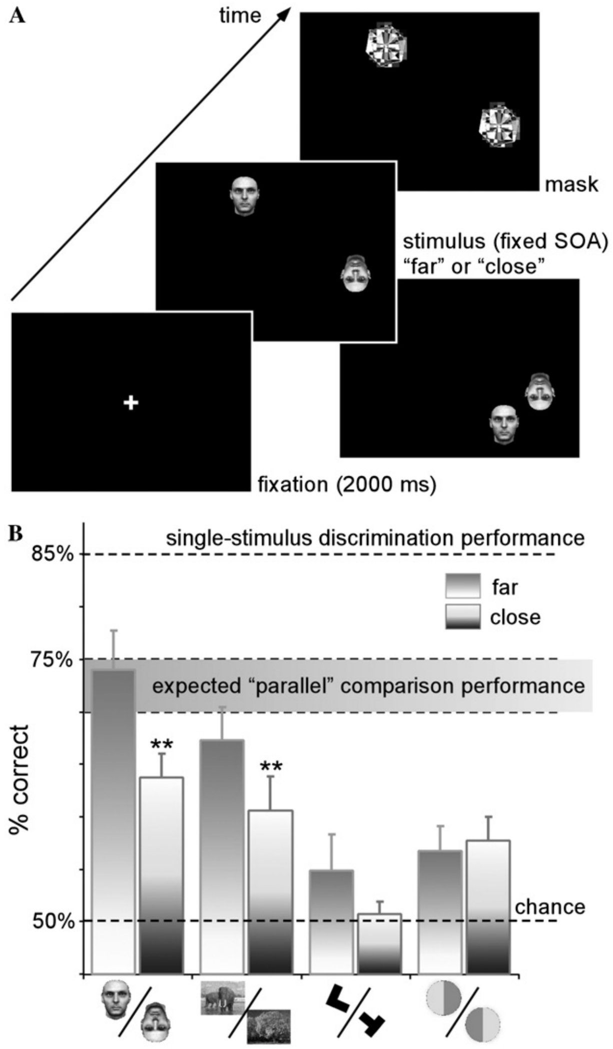

Fig. 4. Comparison task. (A) Experimental protocol. In separate blocks for the four tasks described previously (illustrated here in the case of the face orientation discrimination task), stimuli were presented by pairs and their categories had to be compared (same/different judgment). Stimuli were masked, and the SOA corresponded precisely to that determined to yield 85% correct for the same subject and task in the previous phase (Fig. 3). Performance of this comparison task explicitly requires both stimuli to be correctly identified. Pairs were shown in two possible configurations: ‘‘far’’ (opposite quadrants, at a distance of about 8 deg) or ‘‘close’’ (same quadrant, about 3 deg apart) from each other. Figure not to scale. (B) Results. In the ‘‘far’’ condition, natural stimuli lead to near-optimal comparison performance, while the synthetic stimuli (randomly rotated letters, bisected disks) are poorly compared: these latter stimuli seem to undergo a form of binding problem. In the ‘‘close’’ condition, performance decreases for the natural discrimination tasks, and remains poor for the other tasks. The interference appearing between natural stimuli in this case indicates the presence of a local binding problem.

Fig. 4B presents the average comparison performance obtained in the ‘‘far’’ and ‘‘close’’ conditions for the four discrimination tasks. When two stimuli have to be compared across opposite quadrants (“far” condition), fairly little interference is observed for natural stimuli: upright vs. inverted face discrimination leads to a 74% comparison performance, and animal vs. non-animal scene categorization to a 67.3% comparison performance, close to the “ideal” performance level. On the other hand, the other two tasks (discrimination of rotated Ls and Ts, and of bisected 2-color disks) are performed quite poorly in the same condition (54.9% and 56.7%, respectively). The effect of the task on comparison performance in this “far” condition is significant (one-way ANOVA, $F(3,14) = 7.03$, $p < .005$), and a post hoc test (Tukey–Kramer, $p < .05$) reveals that the natural discriminations (upright vs. inverted faces, animal vs. non-animal scenes) lead to better comparison performance than the other two. This discrepancy confirms many previous dual-task results showing that the first two, natural discrimination tasks suffer much weaker attentional costs than the last two (e.g., Li et al., 2002; Reddy et al., 2004). Note that here the four discrimination tasks are directly comparable, since they have been equated for difficulty (by varying the SOA), leading to 85% correct performance on isolated stimuli. Furthermore, these results are obtained here with no extensive prior training on this comparison task, unlike in previous dual-task studies.

When the two stimuli to be compared are presented within the same quadrant (“close” condition), the strong interference observed for the discrimination of rotated letters or bisected disks remains unaltered (comparison performance is 50.7% and 57.7% correct, respectively; these values are not significantly different from those obtained in the “far” condition; paired t test, $p > .05$). The identification of these stimuli can simply not be performed in parallel, independent of the distance between them. In other words, object binding for these stimuli must rely on a global resource. For the natural discrimination tasks however, even though near-parallel performance had been observed in the “far” condition, interference now appears in the “close” condition. The upright vs. inverted face discrimination and the animal vs. non-animal scene categorization now lead to 63.7% and 60.6% correct comparison performance, respectively. These values are significantly lower than those obtained in the “far” condition (paired t test, $p < .05$). Thus, as predicted, interference for these natural stimuli appears to be spatially specific: strong for neighboring stimuli, almost absent when stimuli are far enough apart. In other words, object binding for these stimuli appears to rely on local mechanisms. Note that this last result is somewhat counterintuitive, since one could reasonably expect that two stimuli should be easier to compare when they are close. This could seem logical, but for the local binding problem that we have demonstrated here.

3.3. Control experiment: Within- vs. between-hemifields comparison

While the present results demonstrate that the binding problem for natural objects and scenes does not operate uniformly over the entire visual space, they do not thoroughly constrain the spatial range of the underlying binding process. In particular, it would be important to know whether the interference observed in the “close” condition depends on competition for an attentional resource that is specific to each cortical hemisphere, or to each quadrant within a given hemisphere. Because so far we have only compared performance between same-quadrant and diagonally opposite-quadrant conditions, there remains the possibility that the visual system might be using a “default binding” strategy within each cortical hemisphere. This would be compatible with proposals that separate attentional resources exist for each hemisphere (Luck, Hillyard, Mangun, & Gazzaniga, 1989; Muller, Malinowski, Gruber, & Hillyard, 2003). To address this possible confound, we replicated the previous comparison experiment (including the first phase used to determine psychometric curves) in four subjects, this time with stimulus pairs constrained to one visual hemifield (randomly determined for each trial).

The results for this control experiment are presented in Fig. 5, and are identical to our previous observations. In the “far” condition, there was a main effect of task (one-way ANOVA, $F(3,11) = 40.6$, $p < .00001$), and post hoc tests (Tukey–Kramer) showed that the natural stimuli (faces, natural scenes) yielded higher (near-optimal) comparison performance than the synthetic ones (randomly rotated letters, bisected disks). In the “close” condition, a decrease (paired t test, $p < .05$) was observed only for the natural scene categorization and face orientation discrimination tasks.

The fact that we could replicate our previous results even when stimuli always have to be compared within a single hemifield suggests that the main factor determining the presence of a binding problem for natural objects and scenes is inter-stimulus distance per se. This effect cannot be predicted by the allocation of independent attentional resources in each hemisphere.

4. Discussion

4.1. Relation to visual search results

Face discrimination, object recognition or scene categorization tasks are generally found to yield serial search slopes in visual search experiments (Brown, Huey, & Findlay, 1997; Nothdurft, 1993; Purcell, Stewart, & Skov, 1996; Rousselet, Thorpe, & Fabre-Thorpe, 2004a, Rousselet, Thorpe, & Fabre-Thorpe, 2004b;

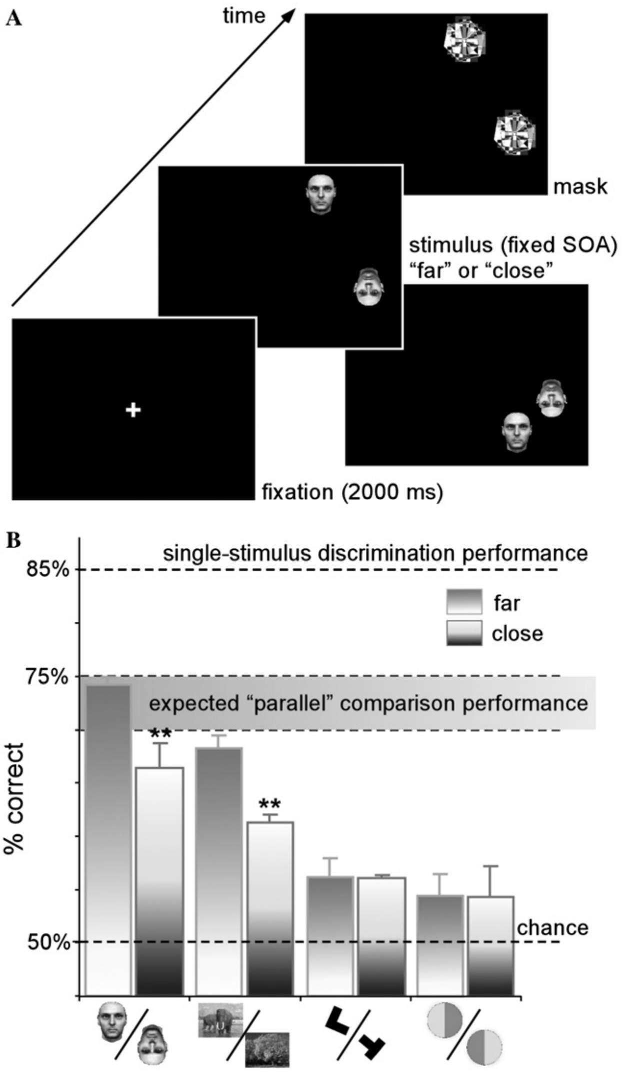

Fig. 5. Control experiment. (A) Experimental protocol. This experiment was designed similarly to the previous one in all respects, except that stimulus pairs in the "far" interference condition were now placed within a single hemifield (randomly determined for each trial). (B) Results. As in the previous case, in the "far" condition comparison performance for the natural discrimination tasks (animal vs. non-animal, upright vs. inverted faces) was near-optimal, while that of the synthetic discrimination tasks (randomly rotated letters, bisected disks) was poor. Furthermore, a significant performance decrease was observed in the "close" condition only for the natural stimuli. This pattern of interference indicates that the observed binding problem remains a purely local problem, even when stimulus positions are restricted to one hemifield.

VanRullen et al., 2004; Wolfe, 1998). Hence it is generally concluded that such processes are not "parallel", and require attention. Our comparison task is somewhat simpler in design but addresses the same question: we show that these processes are in fact "parallel" or "pre-attentive"—but this can only be observed when stimuli are well separated. The crowded displays used in visual search would clearly preclude this conclusion (Reddy et al., 2005).

4.2. Basic features vs. conjunctions

Object binding is often defined as the ability to conjointly recognize object features in a spatially specific manner: if there is a round object and the color green in the same area of space, then it must be an apple. Under this definition, what we have shown here is a true binding process taking place outside the focus of attention. If 'basic' features are defined as those properties that can be processed in parallel, one single basic feature cannot account for our subjects' recognition performance on natural scenes or faces: otherwise interference would not occur either in our 'far' nor 'close' conditions. The most rational way to explain the interference recorded in the "close" condition for these stimuli is thus by assuming that it is due to local binding errors (i.e., local "illusory conjunctions") between features of the neighboring objects. In other words, binding appears to be a local problem for natural scenes and objects.

What would the "features" be that make up these complex natural object categories? This is not a trivial question to answer. It is not obvious that a simple combination of the known "basic" features (color, orientation, curvature, etc.; Wolfe, 1998) could allow one to discriminate natural scenes or faces. If it did, then current artificial object recognition systems, which are often based on such types of features, would be much more efficient than they actually are today. It is more likely, unfortunately, that for efficient object recognition the visual system uses a specific set of features that have not been uncovered yet.

We are open to the possibility that global statistical properties of objects or scenes might participate in this discrimination, as proposed by recent models (Renninger & Malik, 2004; Torralba & Oliva, 2003). Such "histogram-based" models work in fact by detecting "feature conjunctions", i.e., combinations of activity among particular "channels" (e.g., corresponding to the Fourier spectrum). But they are usually blind to spatial relations and are "holistic" (i.e., they work over the entire visual field, just like "default binding"), so they could probably not differentiate between the "far" and "close" interference conditions in our experiments. If such a statistic-based model were to account for our results, it would have to be able to detect these statistical feature conjunctions in a spatially specific manner, i.e., it would have to perform true object binding.

Furthermore, it is worth noting that all so-called "single" or "basic" features are in fact conjunctions, from a computational viewpoint: orientation a conjunction of retinal contrasts with particular respective positions, color a conjunction of specific cone types, etc. The correct distinction might thus need to emphasize the level at which features are bound (low-level binding for orientations vs. high-level binding for complex objects), rather than whether they are "simple features" or "conjunctions".

4.3. Spatial specificity

How spatially specific is the observed process? Here, we compare performance within vs. between quadrants, but what exactly is the spatial range of this type of binding? All we can say is that it is smaller than 8 deg but at least as large as 3 deg (interference was found to occur in the "close" condition, at an inter-stimulus distance of 3 deg, but not in the far condition, at 8 deg) Note that in a dual-task study comparable to our experiment 1, (Fei-Fei et al., 2005) found no strong effect of inter-stimulus distance for distances varying between roughly 4 and 8 deg. More recently, we have reported substantial effects of varying inter-stimulus distance between 0.5 and 3 deg in a visual search paradigm (Reddy et al., 2005). Our present data (control experiment 2) ruled out the possibility that a specific binding process occurs separately in each hemifield (Luck et al., 1989; Muller et al., 2003). We would like to propose instead that the actual spatial range of this binding might have something to do with receptive fields sizes (for an up-to-date review of receptive field sizes in the ventral pathway and their effect on object and scene recognition see Rousselet et al., 2004a).

4.4. Binding and correlated firing

Several researchers have suggested that feature binding in the brain might arise through correlated firing among cells representing various properties of an object, in particular in the gamma frequency band (Eckhorn et al., 1988; Singer & Gray, 1995; von der Malsburg, 1995). Gamma synchrony between neural recording sites is known to decline with increasing cortical distance (Eckhorn et al., 2004; Roelfsema, Lamme, & Spekreijse, 1998) (not taking into account the relation between the encoded representations), which according to the binding-by-synchrony hypothesis, could explain why binding interference ("illusory conjunction") was more frequent for close-by objects in our experiments. This would also be compatible with the observation that the crowding effect in amblyopia (an impairment of object recognition due to close-by interfering objects, similar to the effects described here) is associated with a loss of synchronization between cortical cells (Roelfsema, Konig, Engel, Sireteanu, & Singer, 1994). This explanation is only speculative however, it cannot easily account for the interference observed over much longer distances with artificial geometric shapes. Unfortunately, the neuronal correlates of binding for natural and synthetic objects remain an open issue, which cannot be directly addressed by the current experiments.

4.5. Computational implications

In computational terms, it is easy to see how in a hierarchical neural network (Fukushima & Miyake, 1982; Riesenhuber & Poggio, 1999) neurons at a certain level can be made to respond to conjunctions of features at the preceding level in a spatially specific way. This is actually one definition for the concept of receptive field! It is also straightforward to see that in such a system binding could occur "automatically", i.e., in the absence of attention, but only for well isolated objects (e.g., Mozer & Sitton, 1998). With numerous objects falling inside one receptive field a local "binding problem" arises and competition must be resolved by the effects of attention (Desimone & Duncan, 1995). To summarize, in such a hierarchical (feed-forward) model binding could occur preattentively, interference between stimuli would be observed only locally, and the spatial range of this interference would depend on the size of the relevant receptive fields. This is, in our view, the simplest and most direct explanation of our results.

Hierarchical systems suffer from the combinatorial explosion problem: not all feature combinations of one level can be explicitly bound by neural populations at the next, lest the size of the network turn out to be unreasonably large. To circumvent this problem, the system must select the most relevant objects and categories to be represented. Practice and experience are likely to play a role in this selection, and so it makes sense that familiar, natural object categories such as faces and scenes should be preferred to synthetic, unfamiliar objects such as randomly rotated letters or bisected 2-color disks. These latter objects, unsupported by a hierarchical (i.e., "hardwired") binding, could only be recognized under the effect of attention (even when presented in isolation), as advocated in the original "Feature-Integration Theory" (Treisman & Gelade, 1980). Note that under certain circumstances, some synthetic stimuli can be processed just as efficiently as natural ones, in particular when they happen to be highly familiar: in a recent study Fei-Fei et al. (2005) confirmed that letter recognition (L/T discrimination) in the absence of attention can be made as efficient as natural scene categorization, if the letters are always presented upright (a situation with which every adult subject would be very familiar).

Our results therefore point to two distinct types of "object binding" occurring in the visual system: "hardwired" binding for familiar and natural objects, as in classical hierarchical neural networks; and "arbitrary" (attention-dependent) binding for less familiar or synthetic objects, as proposed by the Feature-Integration Theory. Mapping these processes back onto the visual system, one could propose (as did Lamme & Roelfsema, 2000) that feed-forward selectivities recorded in the ventral pathway could correspond to a "preattentive" hierarchical object recognition system. Refinements of these feed-forward representations under the effect of attention (the "object files" of Treisman) might develop in the same neuronal populations after a sufficient amount of time (Roelfsema et al., 1998; Sugase, Yamane, Ueno, & Kawano, 1999), due to underlying feedback processes, but might also map onto other visual structures altogether (e.g., parietal or prefrontal cortices).

Acknowledgments

The authors thank Christof Koch, Pietro Perona and Guillaume Rousselet for useful discussions, as well as Leila Reddy for helpful comments on the manuscript.

References

Ashbridge, E., Cowey, A., & Wade, D. (1999). Does parietal cortex contribute to feature binding? *Neuropsychologia, 37*(9), 999–1004.

Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: detection and discrimination tasks. *Perception & Psychophysics, 60*(1), 1–23.

Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Perception & Psychophysics, 48*(1), 45–58.

Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: do faces pop out. *Perception, 26*(12), 1555–1570.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., et al. (1988). Coherent oscillations: a mechanism of feature linking in the visual cortex. Multiple electrode and correlation analyses in the cat. *Biological Cybernetics, 60*(2), 121–130.

Eckhorn, R., Gail, A. M., Bruns, A., Gabriel, A., Al-Shaikhli, B., & Saam, M. (2004). Different types of signal coupling in the visual cortex related to neural mechanisms of associative processing and perception. *IEEE Transactions on Neural Networks, 15*(5), 1039–1052.

Fei-Fei, L., VanRullen, R., Koch, C., Perona, P., 2005. Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. Visual Cognition, in press.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science, 269*(5225), 853–855.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition, 15*, 455–469.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neuroscience, 23*(11), 571–579.

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 of the United States of America, 99*(14), 9596–9601.

Luck, S. J., Hillyard, S. A., Mangun, G. R., & Gazzaniga, M. S. (1989). Independent hemispheric attentional systems mediate visual search in split-brain patients. *Nature, 342*(6249), 543–545.

Mozer, M. C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). New York: Psychology Press.

Muller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature, 424*(6946), 309–312.

Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception, 22*(11), 1287–1298.

Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception, 25*(9), 1091–1108.

Reddy, L., VanRullen, R., Koch, C., 2005. Inter-stimulus distance effects in visual search. submitted.

Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision, 4*(2), 106–117.

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research, 44*(19), 2301–2311.

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron, 24*(1), 111–125.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.

Roelfsema, P. R., Konig, P., Engel, A. K., Sireteanu, R., & Singer, W. (1994). Reduced synchronization in the visual cortex of cats with strabismic amblyopia. *European Journal of Neuroscience, 6*(11), 1645–1655.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*(6700), 376–381.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5*(7), 629–630.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004a). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences, 8*(8), 363–370.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004b). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Research, 44*(9), 877–894.

Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences of the United States of America, 99*(16), 10917–10922.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*, 555–586.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*(6747), 869–873.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14*(3), 391–412.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107–141.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual-tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience, 16*(1), 4–14.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology, 5*(4), 520–526.

Wolfe, J. M. (1998). Visual Search. In H. Pashler (Ed.), *Attention* (pp. 13–73). London, UK: University College London Press.