Competition and selection during visual processing of natural scenes and objects

Rufin VanRullen   CNS Program - Division of Biology, California Institute of Technology, Pasadena, CA, USA

Christof Koch   CNS Program - Division of Biology, California Institute of Technology, Pasadena, CA, USA

When a visual scene, containing many discrete objects, is presented to our retinae, only a subset of these objects will be explicitly represented in visual awareness. The number of objects accessing short-term visual memory might be even smaller. Finally, it is not known to what extent "ignored" objects (those that do not enter visual awareness) will be processed –or recognized. By combining free recall, forced-choice recognition and visual priming paradigms for the same natural visual scenes and subjects, we were able to estimate these numbers, and provide insights as to the fate of objects that are not explicitly recognized in a single fixation. When presented for 250 ms with a scene containing 10 distinct objects, human observers can remember up to 4 objects with full confidence, and between 2 and 3 more when forced to guess. Importantly, the objects that the subjects consistently failed to report elicited a significant negative priming effect when presented in a subsequent task, suggesting that their identity was represented in high-level cortical areas of the visual system, before the corresponding neural activity was suppressed during attentional selection. These results shed light on neural mechanisms of attentional competition, and representational capacity at different levels of the human visual system.

Keywords: attention, competition, capacity, negative priming, natural scenes.

Introduction

Every eye fixation brings to our retinae a new visual scene, from which the visual system must extract the most relevant information. Clearly, not all objects from a typical scene will be consciously registered (Rensink et al., 1997; O'Regan et al, 1999; Simons & Levin, 1998). Among those that will, many will not be consolidated into visual memory, and will be rapidly forgotten (Sperling, 1960; Baddeley, 1986). The visual system must therefore continuously and actively select at different stages the properties or objects relevant to current behavior and higher cognitive functions. How does this selection occur? What determines, and what is the relation between what we *see*, what we *almost see*, and what we *fail to see*?

There is increasing evidence that at least some form of high-level representation of the visual scene can be accessed very rapidly (Thorpe et al., 1996; VanRullen & Thorpe, 2001), in an automatic and possibly unconscious way (Ohman & Soares, 1994, 1998; Esteves et al., 1994; Dehaene et al., 1998; Bar et al, 2001; VanRullen & Koch, in press). This representation can be detailed enough to allow subjects to detect an animal in a briefly flashed image, or to categorize a scene in rapid serial visual presentation (RSVP; Potter & Levy, 1969; Potter, 1976; Bar & Biederman, 1998; Coltheart, 1999). In contrast, consciously recognizing an object probably requires some form of attention to be drawn selectively to this object

(Rensink et al., 1997; Simons & Chabris, 1999; Mack & Rock, 1998). Further selection might be required in deciding what objects should be consolidated in memory, and what objects can be forgotten. Figure 1 illustrates this continuous selection process among successive levels of representation.

The capacity of these different levels of visual representation (preconscious, conscious, short-term memory) can be assessed with specific paradigms. Free recall is typically used to access the contents of immediate working memory, in general found to contain around 4 objects (Sperling, 1960; Broadbent, 1975; Pylyshyn & Storm, 1988; Yantis 1992; Luck & Vogel, 1997; Cowan 2001). Implicit measures, such as performance in forced-choice recognition, or visual priming, can be used to determine which objects were perceived, even when they are not explicitly remembered (e.g. Biederman & Cooper, 1991; Bar & Biederman, 1998, 1999).

However, estimates of capacity obtained by different studies with different paradigms, and at different levels of representation, are very unlikely to be comparable. Here we apply a combination of three such paradigms (free recall, forced-choice recognition, and visual priming) on the same complex natural scenes and for the same subjects. Immediately after a large natural scene containing 10 different objects was briefly presented, subjects had to report the objects that they had perceived. They could also "guess" an additional number of objects. Subsequently, these same objects were presented among

other unfamiliar ones in a word-picture matching task. Reaction times were analyzed to reveal visual priming. Surprisingly, the objects that the subjects could neither explicitly report nor guess elicited a significant negative priming effect, suggesting that they had been suppressed at a rather late stage of visual processing.
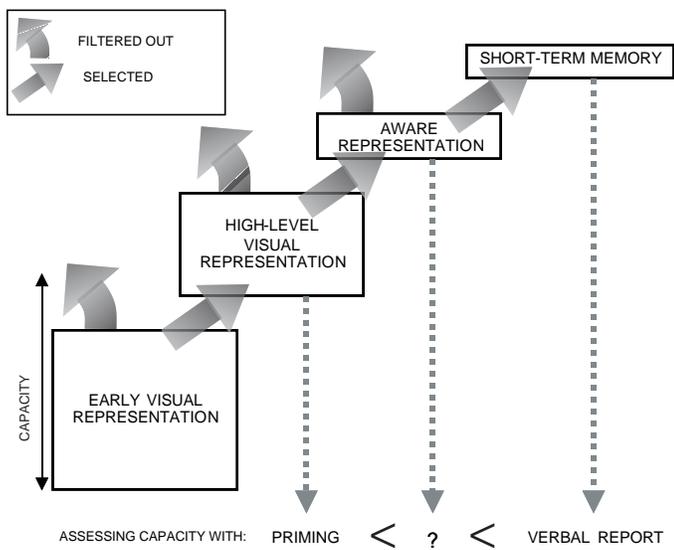

Figure 1. Different levels of representation in the visual system (schematic). At each stage, information can be filtered out or selected to access the following stage. Different experimental paradigms can query the contents of these representations. Verbal report (or free recall) is typically used to estimate the capacity of visual short-term memory. Note that a number of studies refer to visual short-term memory as an early visual buffer, not necessarily conscious (Phillips & Baddeley, 1971; Phillips, 1974; Tulving & Schacter, 1990; Jiang et al, 2000; Magnussen, 2000). Here we adopt a more intuitive definition: an item (object or property of an object) is considered as being stored in short-term memory if it can be recalled, i.e. explicitly reported. In this context, memorized objects necessarily are or have been represented into visual awareness, at the time they are reported. In contrast, implicit measures such as visual priming or performance in forced-choice recognition can determine which objects have reached a high-level representation. Among these objects, some will be selected by attention to enter visual awareness, and a certain number might be filtered out.

It is necessary to stress that negative priming has been known for over 20 years as a reflection of active attentional suppression of ignored objects (Neill 1977; Tipper 1985; Fox 1995). However, it is typically observed in situations where a unique target (attended) object competes with another unique overlapping distractor (ignored) object, and the to-be-attended property (e.g. color) is defined in advance. Here negative priming is reported under "realistic" conditions of stimulation, where different objects of a natural scene compete for

attentional resources and selection, and observers have no a priori bias as to what object or property they should attend to.

Methods

Free Recognition and Forced-Choice Recognition

Each of 10 stimulus scenes (Figure 2A), containing 10 objects, was presented for 250 ms, immediately followed by a strong contrast color mask (a situation designed to approximate an average single fixation). The mask was obtained by superimposing many different samples of white noise that were band-pass filtered at particular spatial frequencies, so that the resulting mask would display a power spectrum resembling that of natural images (i.e. $1/f$). The scene and mask subtended 16 degrees of visual angle in width. Immediately after each scene, subjects were presented with a list of 20 object names, including the 10 target objects. Distractor object names were carefully chosen so that they could have normally been present in the context of the scene. Subjects were asked to report the objects that they had consciously perceived with full confidence (free recognition). After signaling that they were not confident anymore, they had to select a further number of objects (forced-choice recognition), so that the overall number of selected objects, including the ones reported with full confidence, was exactly 10. Note that the term "forced-choice recognition" generally refers to a situation where the number of alternatives is determined by the experimenter. In our case, the number of alternatives is determined by the subjects' performance in the previous "free recognition" task.

Correction for Guessing

R^*, the corrected number of objects reported in free recognition (correct reports that can not be explained by chance), is defined as:

$$R^* = R^+ - R^-$$

where R^+ and R^- are the number of target and distractor objects reported by a given subject for a given scene. Note that high-threshold models (commonly used to estimate capacity; Pashler, 1988; Luck & Vogel, 1997) suggest a slightly different correction method:

$$R^* = (R^+ - R^-) / (1 - R^-/10) .$$

However, R^- is small enough in our case, and the difference between these 2 methods can be neglected.

In the forced-choice recognition part of the report task, the a priori distributions of probability for target and distractor objects are not equal but depend on the previous responses (R^+, R^-) of each subject for each scene. Therefore, the number G^* of "above-chance guesses" can be defined as:

$$\frac{G^+ - G^*}{10 - R^+ - G^*} = \frac{G^-}{10 - R^-}$$

$$\Leftrightarrow G^* = \frac{G^+\left(10 - R^-\right) - G^-\left(10 - R^+\right)}{10 - R^- - G^-}$$

where G^+ and G^- are the number of correct and incorrect guesses.
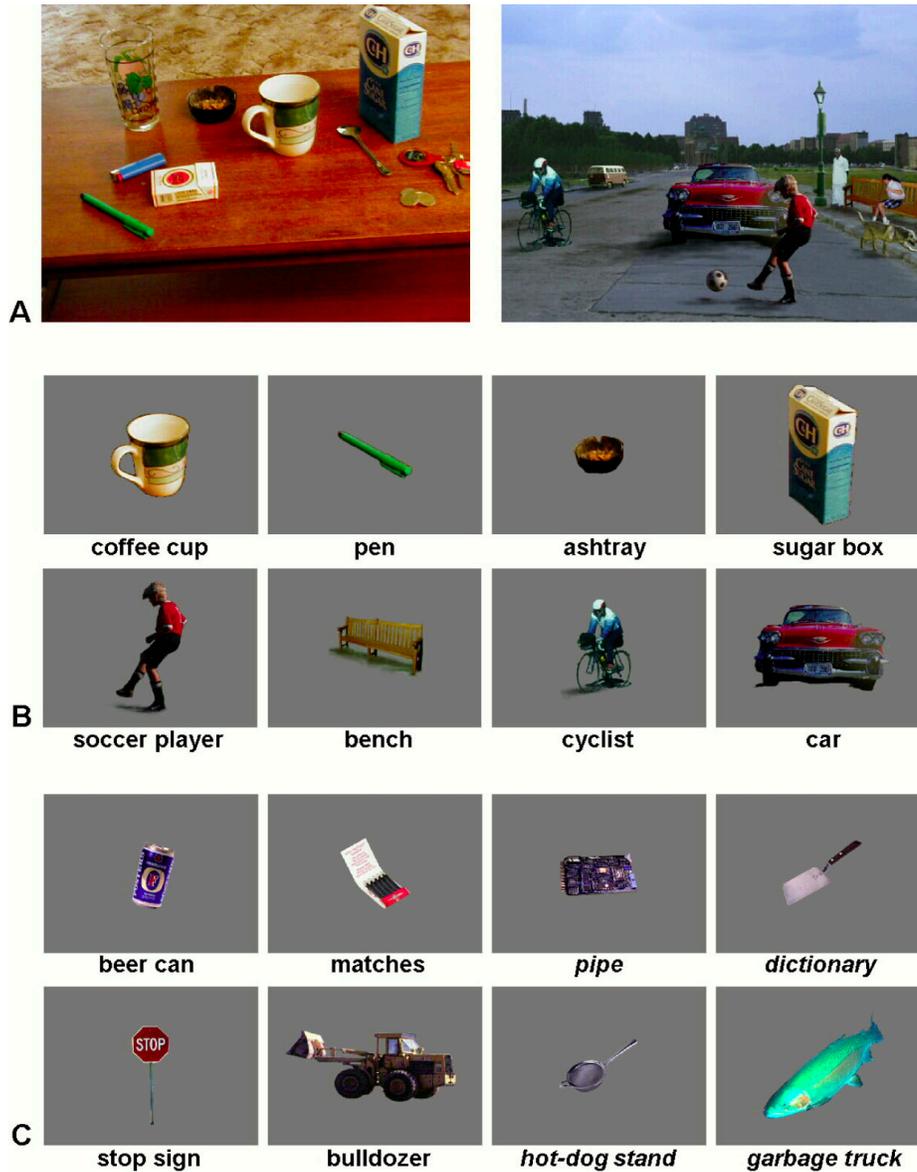
Figure 2. A. Two examples of natural scenes used. All 10 scenes used in this experiment can be viewed at http://www.klab.caltech.edu/~rufin/capacity.html B. Examples of target objects, extracted from these scenes, and presented during the word-picture matching task. The name associated with each object is shown under each image ("match" trials). C. Examples of distractor objects presented in the word-picture matching task. These objects were not present in the scenes used in the report task (free and forced-choice recognition), but the names associated with these objects were part of the list of 20 objects from which subjects had to pick the objects that they had perceived (first row corresponding to the "coffee table" scene; second row corresponding to the "street" scene). Note that in each row, the 2 word-picture pairs on the right correspond to "non-match" trials. Finally, there was also for each scene a set of "new" word-picture pairs, in which neither the object image nor the name had been presented in the previous report task (not shown).

Visual Priming

Immediately after the forced-choice recognition task for each scene, the subjects had to perform a block of 40 trials of a word-picture matching task, in which some of the stimuli were target objects that had appeared in the previous scene. In each trial, an object name was presented for 500 ms and, following an inter-stimulus interval of 1 second, an object image was flashed at fixation for 250 ms. The subjects held down the mouse button continuously, and had to release it as fast as possible, within 1 second, if and only if the object picture matched the previous word. The objects were presented on a uniform grey background of the same luminance as the rest of the screen. Object size was variable, between approximately 2 and 10 degrees of visual angle. The objects that were extracted from the scene were always presented with their original size, at the fixation point. The average change in eccentricity for a given object between its presentation in the scene and its presentation in isolation was around 3.5 degrees. All 20 object names from the free recognition and forced-choice recognition tasks (10 targets and 10 distractors) were presented in this block. The 10 target objects always matched the target names ("match" trials). Five of the 10 distractor words were paired with a matching object, and five with a non-matching object. Finally, in an additional 20 trials of the same block, both the object name and the object picture were totally new (15 "match" and 5 "non-match" trials). Note that the familiarity of the written name (i.e. whether it belonged or not to the list of 20 objects in the previous task) did not predict the status (match/non-match) of the following object, since in both cases the probability of a match trial was 75%. The order of the trials was randomized in each block. Reaction times (RT) were recorded for each trial, and were used as a measure of visual priming.

Experimental Setup

Ten subjects in each group (test and control) participated in the experiment. They were seated in a dark room, 120 cm from a computer screen connected to a SGI (O2) workstation. They were first trained on 2 examples of simple scenes and the corresponding word-picture matching task blocks. The group of control subjects performed the experiment in reverse order, viewing the word-picture matching task before they were presented with each scene and had to report their contents. The reaction times from these subjects in the word-picture matching task were used as a reference (no priming). Furthermore, their performance in the report task (free and forced-choice recognition) allows us to determine if and how object recognition is facilitated by a prior single exposure to target objects.

To summarize, the test subjects were presented with a scene, asked to report (or guess) its contents, then performed the corresponding word-picture matching task; conversely, the control subjects were first asked to perform this word-picture matching task, then viewed the scene, and finally reported its contents. This sequence was repeated 10 times for each group.

Results

Free Recognition

On average, subjects explicitly report 2.28 objects per scene (corrected for guessing; see Methods and Table 1). This number is dependent upon the particular scene, and upon individual subjects. The number of reported objects varies between 1.7 and 3 for different scenes (averaged across subjects), and between 1.8 and 2.7 for different subjects (averaged across scenes).

Table 1. Average Number of Objects Selected in Each Scene.

Objects/ Scene		Correct	Incorrect	Corrected	d'
Free recognition	Test	2.61 (/10) 0.44	0.33 (/10) 0.26	2.28 0.35	1.16 0.22
	Control	3.52 (/10) 0.74	0.21 (/10) 0.15	3.31 0.65	1.63 0.22
Forced-choice recognition	Test	3.96 (/7.39) 0.64	3.08 (/9.67) 0.53	2.28 1.18	0.56 0.30
	Control	3.72 (/6.48) 0.67	2.56 (/9.79) 0.38	2.72 0.71	0.82 0.21

Average number of objects selected in each scene, during the free recognition and the forced-choice recognition tasks, for test and control subjects. The number of remaining elements to choose from is indicated in parenthesis where applicable. Correction for guessing is calculated as described in the Methods. Standard deviation is indicated below each number. d' is also provided for information.

The group of control subjects, who have been presented once with the target objects, performs reliably better (paired t-test, d.f.=9, t=5.55, p<.001). On average, these subjects report 3.31 objects per scene (corresponding to a 45% increase in recognition performance). This increase is paralleled by a corresponding increase of about 40% of the d'. Here again, performance varies across individual subjects (from 2.7 to 4.8) and scenes (from 2.1 to 4.2). Interestingly, the number of errors (R) is not higher for these control subjects than for the test subjects (0.21 errors per scene

versus 0.33 errors per scene), indicating that this improvement truly reflects a facilitation of object recognition, and not simply a higher degree of confidence, or a change in report strategy.

Forced-Choice Recognition

The average (corrected) number of correct "guesses" for the group of test subjects is 2.28 (see Table 1). This number varies between 0.46 and 4.29 for individual scenes (averaged across subjects), and between 0 and 3.68 for individual subjects (averaged across scenes). For control subjects, the average number of correct guesses is 2.72, ranging from 1.1 to 4.5 for individual scenes and from 1.67 to 3.5 for individual subjects. Because control subjects had already reported more correct objects than test subjects in the free recognition task, they had fewer target and more distractor objects to choose from in the forced-choice recognition task. Taking into account these a priori probabilities for each group, this corresponds to a 36% increase in recognition probability for control subjects versus test subjects. Note that the d' measure also parallels this increase of about 40% (Table 1).

Figure 3 presents the combined results from the free recognition and forced-choice recognition tasks for each of the 10 scenes that were used as stimuli. The number of objects correctly "perceived" by test subjects (i.e., either explicitly reported, or guessed in the forced-choice recognition paradigm) varies between 2.3 and 6.1. After a single prior exposure to target objects, control subjects correctly perceive between 4.1 and 7.5 objects per scene.

It is not entirely clear how many of these objects have reached a conscious level of representation. A lower bound of around 4 objects can be recalled from visual short-term memory. This number is compatible with previous measurements of the capacity of short-term memory, generally believed to contain between 4 and 6 individual items (Sperling, 1960; Broadbent, 1975; Pylyshyn & Storm, 1988; Yantis 1992; Cowan 2001). Among the remaining objects, a certain number (and possibly all) might have accessed visual awareness, but without leaving a strong enough trace for later recall.

It is important to note that the number of objects perceived can depend on the particular scene presented, and probably on specific properties of each target object, such as its overall saliency. Among the factors that might determine whether an object will be reported or not, retinal eccentricity (that is, distance from fixation point) and size seem to be of particular importance. As compared to an average over all objects, the distance from fixation point is 15% smaller (t-test, d.f.=9, t=5.39, p<.001) for the objects reported by test subjects, and 11% smaller (t=4.25, p<.005) for those reported by control subjects. The size of the objects reported by test subjects is also 25% larger (t>10, p<.0001) than the average size of all objects, and the objects reported by control subjects are 22% larger (t>10, p<.0001). Finally, the objects that
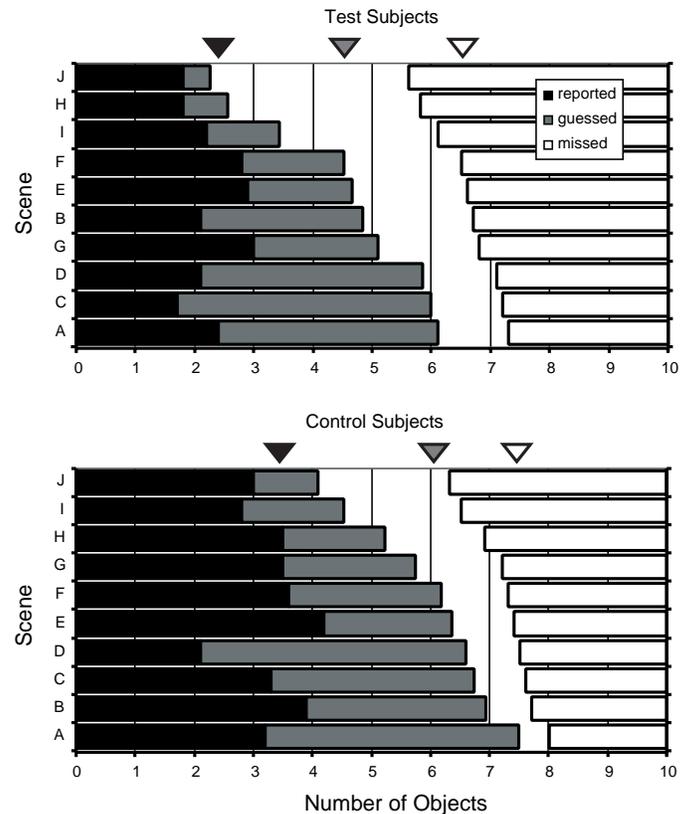

Figure 3. Average number of objects correctly reported, guessed, or missed for each scene, and for the 2 subject groups. In each panel, the scenes are ordered according to the sum of the number of reported and guessed objects. The scene labels (from A to J) reflect this order for control subjects. The "coffee table" and "street" scenes from figure 2 correspond to labels C and G, respectively. The numbers of reported and guessed objects have been corrected for chance guessing as described in the Methods. This correction explains why a certain number of objects in each scene are not assigned to any category: they correspond to correct responses that were discarded by this correction. The triangles indicate the average numbers of reported, guessed and missed objects for each subject group.

were guessed during forced-choice recognition show a significant (t>4, p≤.001) trend in the other direction, being 8% smaller for test and 10% smaller for control subjects than the average size for all objects. Conversely, the objects that were missed (i.e., neither explicitly reported, nor guessed during forced-choice recognition) are 17% smaller than average (t=7.15, p<.0001) for test subjects and 20% smaller (t>10, p<.0001) for controls, while their distance from fixation is roughly 6% higher than the average (although this number is only significant for test subjects, at the p<.001 level, t=5.32).

To summarize the results described so far, up to 7.5 objects from a complex natural scene can be identified in a single fixation, although 6 would be a more reliable

(and conservative) estimate. Up to 4 of these objects can be consolidated into visual short-term memory and are reported by subjects with high confidence as having been "seen". We now turn to the question of the remaining objects, those that were neither reported in free recognition nor guessed in forced-choice recognition (the "missed" objects). Whereas these objects obviously did not access a conscious level of representation, it is still possible that they could have reached some "high" level of representation, i.e., been recognized before being filtered out. In other words, does the observed limitation occur at the level of visual awareness or visual short-term memory, or is this limitation a consequence of a low-level selection, occurring earlier on in the visual system?

Visual Priming

When a particular stimulus (hereafter called the "prime") is presented to the visual system, even under conditions where it is not consciously perceived or remembered, it elicits a specific trace of neural activity, that can modify the processing of a subsequent repetition of the same stimulus (hereafter the "probe"). This phenomenon, known as visual priming, can take two distinct forms: either a stimulus-specific facilitation (Biederman & Cooper, 1991; Bar & Biederman, 1998, 1999), or a stimulus-specific impairment of subsequent visual processing (Neill, 1977; Tipper, 1985). While the former effect (positive priming) usually occurs for the objects that are selected by visual attention (or under conditions of low attentional load), the latter (negative priming) is generally thought to reflect the suppression of ignored objects during attentional selection (e.g. Tipper & Driver, 1988; Fox, 1995; Moore, 1996), although alternative theories have been proposed (Neill et al, 1992; Park & Kanwisher, 1994). Visual priming has been shown to be invariant to low-level picture manipulations (translation, reflection; Biederman & Cooper, 1991), and specific to higher-level properties of the stimulus, such as its semantic category (Allport et al, 1985; Tipper & Driver, 1988).

In order to determine whether objects of a particular group (e.g., missed objects) were perceived when the scene was presented, a block of 40 trials of a word-picture go/no-go matching task was performed after each entire report sequence (i.e. only once, after both free and forced-choice recognition were completed for a scene). The target objects from the previous scene were extracted from their background and presented in this task, among other trials containing "new" objects that had not been present in the scene. On average, the delay between the presentation of the whole scene and the presentation of one of these 40 word-picture matching trials was around 2 minutes, that is, well under the reported duration of visual priming (Bar & Biederman, 1998; DeSchepper & Treisman, 1996).

We reasoned that if an object was positively (resp. negatively) primed, the actual reaction time should be shorter (resp. longer) than the reaction time of a control subject, viewing the same object for the first time. In order to make reaction times comparable between the test and control subject groups, we normalized the RTs of each test subject so that their mean and standard deviation for the set of new objects would match the mean and standard deviation of RTs of control subjects on these new objects. We then compared the RT obtained for each target object (i.e., an object that was present in the original scene) to the median RT of control subjects on the same object (in other words, this median RT was considered as a reference). If there was no significant priming effect, on average 50% of the RTs would fall below this reference, and 50% above (since there could have been no priming for the control subjects group). This is what we observed for the set of objects that were guessed in the forced-choice recognition task: 49% of these objects elicited RTs below the reference, and this proportion was not significantly different from 50% (χ^2 test, 396 observations, d.f.=1, χ^2=.09, p=0.8). On the other hand, 55% of the objects that were explicitly reported in the free recognition task elicited RTs that were shorter than the reference, suggesting a non-significant (261 observations, χ^2=2.39, p=.1) positive priming effect, whereas 57.5% of the RTs on missed objects were longer than the reference, indicating a significant (343 observations, χ^2=7.58, p=.005) negative priming effect for these objects. Whereas the former effect (positive priming) can be naturally expected to occur for objects that the subjects explicitly reported (because these objects have obviously been identified), the latter effect is more surprising. Indeed, when a subject reliably fails to report certain objects from the scene, it would be rather intuitive to conclude that these objects were not perceived. However, the negative priming effect suggests that these objects were in fact represented in the visual system, but that this representation was eventually suppressed.

This negative priming effect is also significant when comparing mean RT (paired t-test, t(9)=3.27, p=.01) and error rate (t(9)=3, p=.015) between the set of missed objects and the set of new objects (Figure 4). These latter effects are not significant (t(9)=2.2, p=.055 for RTs; t(9)=1.48, p=.17 for error rates) for the group of control subjects, indicating again that the priming effects are indeed due to the prior perception of target objects in the scene. Additionally, the magnitude of this negative priming (calculated as the difference between error rates for "missed" vs. "new" objects) was stronger for test than control subjects (t(9)=2.96, p=.016). This effect is in fact strong enough (and in particular, stronger than the positive priming observed for explicitly reported objects) to be observed when we average over the entire set of target objects (whether explicitly reported, guessed, or missed): the overall error rate in the word-picture

matching task is significantly (paired t-test, t(9)=2.4, p=.04) higher for target objects (6.2%) than for "new" objects that do not belong to the original scenes (4.0%). Once again, this comparison is not significant for control subjects (t(9)=1.13, p=.29).

This observation is particularly important because it rules out alternative explanations based on the correlational nature of our analysis. Indeed, our subjects select by their performance which objects belong to the class of reported, guessed or missed objects for which priming will later be tested. One could therefore argue that our analysis only reveals correlations between bad performance in both the report task and the reaction time
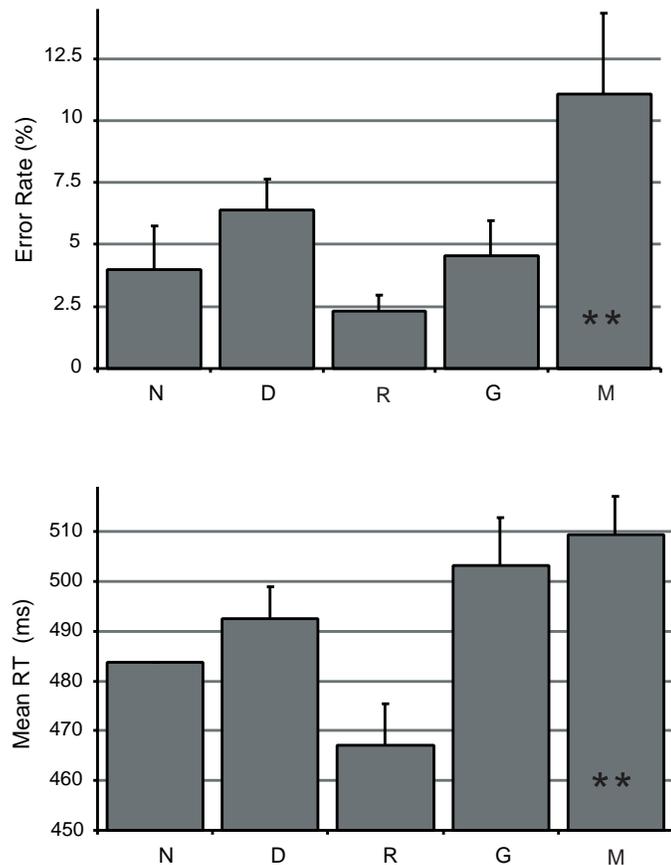

Figure 4. Mean error rates (top) and reaction times (bottom) in the word-picture matching task. Trials are grouped according to the performance of the subject in the previous report task: a target object can be either explicitly reported (R), guessed in forced-choice recognition (G), or missed (M). New trials (N) indicate that the object was not present in the previous scene, nor in the list of 20 object names. Distractor trials (D) refer to object names that were present in the list, but not in the previous scene. Error bars reflect standard error of the mean. The normalization procedure described in the Methods section implies that across-subjects variance of reaction times is zero for "New" objects. Performance for each trial group was compared to performance on new trials (paired t-test, d.f.=9). The star symbols indicate significance at the p≤.01 level.

task. However, this is not true in our case because the group of test subjects actually performs worse on the overall set of target objects, independent of the correlation among images drawn from these three categories.

One could also argue that subjects could recognize written names as part of the previous list, and use this information to bias their response in the word-picture matching task. In that case, the same "negative priming" should also be observed for "distractor" names, those that were actually presented in the previous list but not in the scene (indeed, from the subject's point of view, there is no way to tell these objects from the "missed" objects). However, reaction times obtained for these distractor objects in the priming task are significantly shorter (paired t-test, t(9)=2.55, p=.03) than the ones for "missed" objects, and the error rates significantly lower (paired t-test, t(9)=2.49, p=.035). These RTs and error rates for "distractor" objects are not significantly different (t(9)=1.37, p=.2 for RTs; t(9)=.12, p=.9) from those obtained for "new" objects . In other words, the fact that a name is recognized as part of the previous list, but not part of the scene, cannot by itself account for the observed negative priming.

Yet another possible interpretation of this result could be that the difference between test and control subjects arises from a form of interference between the two tasks. For example, when presented with a missed object in the word-picture matching task, a test subject could realize that he (or she) failed to report this object as part of the previous scene. This in turn might interfere with the generation of the motor response. There could be no such effect for control subjects, who have not yet viewed the scene at the time of the word-picture matching task. However, because such an error judgment would require not only the identification of the object, but also access to the memory of responses from the previous task, one would expect it to mostly affect the longest RTs, i.e., those for which the subject has enough time to make this sort of judgment. In contrast, the shortest RTs would most probably reflect an automatic object recognition process. We find that the probability of generating a motor response for a missed object before 400 ms post-stimulus is already significantly (paired t-test, d.f.=9, t=4.15, p<.005) smaller than the probability of responding to a new object (15% in the former case versus 26% in the latter), suggesting that object recognition itself, and not (only) later cognitive judgments, is impaired in the case of missed objects. In other words, this impairment is certainly a true negative priming effect, indicating that missed objects from the scene have indeed accessed a high level of representation, even if the resulting neural activity was too weak, or did not last long enough, to allow these objects to be consciously reported.

Discussion

When a novel natural visual scene is presented to our retinae, we almost immediately and automatically extract its overall meaning, its "gist" (Wolfe, 1998). In addition, a certain number of individual objects usually complement this representation. When asked to describe what these objects are, observers will usually report 2 or 3 objects with confidence. If they have been exposed to the target objects shortly before, they will most likely be able to report around 4 objects. Even without full confidence, if forced to choose from a list of possible objects, observers can select the correct objects well above chance. This brings the total number of perceived objects up to 6, although some of them might not be explicitly remembered. Prior exposure to the target objects can even increase this total to almost 8 objects. How many of these objects are represented in visual awareness remains unclear, but this number is certainly greater than 4, since in many cases 4 objects or more are explicitly remembered by the observer. Finally, a subject will completely *fail* to report between 2 and 4 out of 10 objects, depending on the particular scene. Note that, for such a failure to occur, the subject must judge other distractor objects more likely to have been present in the scene. In other words, the observer must be confident to a certain degree that they have *not* perceived the target objects in the scene. However, when viewing these same objects in a following task, the subject will tend to respond slower and make more mistakes than for a set of completely new objects (negative priming). Therefore, these objects must have been processed to a certain extent by the visual system, before being filtered out.

This sequence of selection among different levels of representation can be better understood in terms of the underlying neural mechanisms. The early representation that is mediated by neural populations in striate and early extrastriate visual areas (i.e., V1, V2...) most probably describes the scene in a spatially uniform way, except for an enhanced resolution towards the center of the visual field, and a degradation towards the periphery, due to retinal and cortical magnification factors. The competition taking place between neurons at this level is unlikely to account for object-based selection, since the receptive fields will in general be too small, and the selectivities too coarse, to allow the representation of individual objects. In consequence, most if not all of the objects present in the visual scene will be represented (at least partly and/or temporarily) at the level of V4 and in its postsynaptic target areas in the inferior temporal cortex (IT) and in the equivalent regions of the human temporal lobe (e.g., fusiform gyrus), where neural populations as well as individual neurons have been found to code specifically for certain object categories such as faces, houses or chairs (Allison et al, 1999; Aguirre et al, 1998; Epstein and Kanwisher, 1998; Ishai et al, 1999; Chao et al, 1999). A recent electrophysiological study in the macaque by Sheinberg and Logothetis (2001) indicates that objects in natural cluttered scenes such as the ones used here can activate selective neurons in infero-temporal cortex in a manner very similar to an isolated presentation of the same objects. There is supportive experimental evidence that some degree of object-based competition within and between neurons takes place at this level. For example, 2 objects falling inside the same neuronal receptive field are known to compete for attentional resources in order to dominate the neuronal response (Moran & Desimone, 1985; Desimone & Duncan, 1995; Reynolds et al, 1999). As a result of this competition, a certain number of objects (around 4.5 or more in light of the present results) will be selected to receive attentional resources, while the representation of the remaining objects (between 2 and 4 in a scene containing 10 objects) will be *actively inhibited*, so as to avoid interference.

Neurons coding for "ignored" objects will not participate in the following stages of this sequence of processing. However, because they are not passively but actively suppressed or inhibited (either in IT or its post-synaptic targets), the neural activity resulting from a subsequent presentation of the same object will first need to overcome the long-lasting effects of this suppression before the neurons can be made to respond again. This might constitute the neural basis of the negative priming phenomenon (Tipper, 1985). What is remarkable here from a biophysical point of view is that a single exposure of an image, with an associated neural activity most likely lasting less than one second in duration (Kreiman et al., 2000) must give rise to some sort of long-lasting synaptic effect that can lead to a less effective neural representation many minutes later when the same image is flashed on again.

Similarly, a single prior presentation of a target object in isolation (such as when the control subjects performed the word-picture matching task before viewing the scene) will trigger some sort of facilitation in the neurons coding specifically for this object, that can last long enough to enhance later selection of this object, when presented in the context of the scene. This corresponds to a positive priming effect. The number of selected objects can be enhanced in such a way (approximately from 4.5 to more than 6), suggesting that the capacity limitation at this level is not a "hard" limitation, but one that can be overcome in particular situations.

It is striking to notice that the negative priming effect obtained here can be much stronger than the corresponding positive priming observed for selected objects. Indeed, the "net" effect observed on all target objects (whether correctly reported, guessed or rejected) is a significant negative priming one. In contrast, most psychophysical studies more readily appear to observe positive priming (e.g. Biederman & Cooper, 1991; Bar & Biederman, 1998). This discrepancy might arise from the

fact that in our case, the "prime" stimulus is not presented in isolation, but in a cluttered scene containing many objects. This might force the visual system to activate attentional selection mechanisms, inhibiting the representation of certain objects which would otherwise (if presented in isolation) receive full attentional resources. By comparison, other studies do not in general require the visual system to actively select among many different competing stimuli.

What is the fate of the objects whose representation survived the competition at this stage? How do they finally give rise to an explicit percept or a vivid memory? The first obvious conclusion is that the activity of the neural populations representing these objects will last longer or be stronger in some manner than for "ignored" objects, because in the case of selected objects, neural activity has been facilitated, while it was suppressed for the other objects. It is possible that a prolonged duration of neural activity in the inferior temporal cortex could be a necessary condition for the neural correlates of visual consciousness (Crick & Koch, 1990; Subramaniam et al, 2000; Bar et al, 2001). If this hypothesis were true, our results suggest that a possible role for such an integration period could be to leave enough time for competition between stimuli to be resolved, before the selected objects can enter visual awareness. Alternatively, this information could be transmitted to higher-level neuronal areas, such as parahippocampal structures or prefontal areas, which receive direct connections from inferotemporal cortex and whose neurons are selective to visual categories (Distler et al, 1993; Suzuki, 1996; Miyashita & Hayashi, 2000; Kreiman et al., 2000). Such areas could then mediate awareness and/or working memory of the objects selected by visual attention (Suzuki, 1999; Crick & Koch, 1998). The present results however do not allow to differentiate between these two hypotheses.

Acknowledgments

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron, 21*(2), 373-383. [PubMed]

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex, 9*(5), 415-430. [PubMed]

Allport, D. A., Tipper, S. P., & Chmiel, N. R. J. (1985). Perceptual integration and postcategorical filtering. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI* (pp. 107-132). Hillsdale, NJ: Erlbaum.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Bar, M., & Biederman, I. (1998). Subliminal visual priming. *Psychological Science, 9*(6), 464-469.

Bar, M., & Biederman, I. (1999). Localizing the cortical region mediating visual awareness of object identity. *Proceedings of the National Academy of Science USA, 96*(4), 1790-1793. [PubMed]

Bar, M., Tootell, R.B., Schacter, D.L., Greve, D.N., Fischl, B., Mendola, J.D., Rosen, B.R., & Dale, A.M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron, 29*(2), 529-535. [PubMed]

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20*(5), 585-593. [PubMed]

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3-18) Wiley.

Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *Neuroreport, 10*(14), 2945-2950. [PubMed]

Coltheart, V. (Ed.). (1999). *Fleeting memories: cognition of brief visual stimuli*. Cambridge: MIT Press.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1). [PubMed]

Crick, F., & Koch, C. (1990). Some reflections on visual awareness. *Cold Spring Harbor Symposium on Quantitative Biology, 55*, 953-962. [PubMed]

Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex, 8*(2), 97-107. [PubMed]

Dehaene, S., Naccache, L., Le Clec'H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature, 395*(6702), 597-600. [PubMed]

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: implicit coding without attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22*(1), 27-47. [PubMed]

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

Distler, C., Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. *Journal of Comparative Neurology, 334*(1), 125-150. [PubMed]

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601. [PubMed]

Esteves, F., Parra, C., Dimberg, U., & Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology, 31*(4), 375-385. [PubMed]

Fox, E. (1995). Negative priming from ignored distractors in visual selection- a review. *Psychonomic Bulletin, 2*(2), 145-173.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., & Haxby, J.V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Science USA, 96*(16), 9379-9384. [PubMed]

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*(3), 683-702. [PubMed]

Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience, 3*(9), 946-953. [PubMed]

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281. [PubMed]

Mack, A., & Rock, I. (1998). *Inattentional Blindness.* Cambridge MA: MIT Press.

Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neuroscience, 23*(6), 247-251. [PubMed]

Miyashita, Y., & Hayashi, T. (2000). Neural representation of visual objects: encoding and top-down activation. *Current Opinion in Neurobiology, 10*(2), 187-194. [PubMed]

Moore, C. M. (1996). Does negative priming imply preselective identification of irrelevant stimuli? *Psychonomic Bulletin and Review, 3*(1), 91-94.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782-784. [PubMed]

Neill, W. T. (1977). Inhibition and facilitation processes in selective attention. *Journal of Experimental Psychology. Human Perception and Performance, 3*, 444-450.

Neill, W.T., Valdes, L.A., Terry, K.M., & Gorfein, D.S. (1992). Persistence of negative priming: II. Evidence for episodic trace retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 18*(5), 993-1000. [PubMed]

Ohman, A., & Soares, J. J. (1994). "Unconscious anxiety": phobic responses to masked stimuli. *Journal of Abnormal Psychology, 103*(2), 231-240. [PubMed]

Ohman, A., & Soares, J. J. (1998). Emotional conditioning to masked stimuli: expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. *Journal of Experimental Psychology. General, 127*(1), 69-82. [PubMed]

O'Regan, J.K., Rensink, R.A., & Clark, J.J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature, 398*(6722), 34. [PubMed]

Park, J., & Kanwisher, N. (1994). Negative priming for spatial locations: identity mismatching, not distractor inhibition. *Journal of Experimental Psychology. Human Perception and Performance, 20*(3), 613-623. [PubMed]

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics, 44*(4), 369-378. [PubMed]

Phillips, W.A., & Baddeley, A.D. (1971). Reaction time and short-term visual memory. *Psychonomic Science, 22*, 73-74.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics, 16*, 283-290.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*(1), 10-15. [PubMed]

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning & Memory, 2*(5), 509-522. [PubMed]

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179-197. [PubMed]

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368-373.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19*(5), 1736-1753. [PubMed]

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *Journal of Neuroscience, 21*(4), 1340-1350. [PubMed]

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review, 5*(4), 644-649.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074. [PubMed]

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(498), (whole issue).

Subramaniam, S., Biederman, I., & Madigan, S. (2000). Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences. *Visual Cognition, 7*(4), 511-535.

Suzuki, W. A. (1996). The anatomy, physiology and functions of the perirhinal cortex. *Current Opinion in Neurobiology, 6*(2), 179-186. [PubMed]

Suzuki, W. A. (1999). The long and the short of it: memory signals in the medial temporal lobe. *Neuron, 24*(2), 295-298. [PubMed]

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

Tipper, S. P. (1985). The negative priming effect: inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology A, 37*(4), 571-590. [PubMed]

Tipper, S. P., & Driver, J. (1988). Negative priming between pictures and words in a selective attention task: evidence for semantic processing of ignored stimuli. *Memory and Cognition, 16*(1), 64-70. [PubMed]

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*(4940), 301-306. [PubMed]

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience, 13*(4), 454-461. [PubMed]

VanRullen, R., & Koch, C. (in press). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience*.

Wolfe, J. M. (1998). Visual memory: what do you know about what you saw? *Current Biology, 8*(9), R303-304. [PubMed]

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology, 24*(3), 295-340. [PubMed]