Visual Search and Dual Tasks Reveal Two Distinct Attentional Resources

Rufin VanRullen[1],*, Lavanya Reddy[2],*, and Christof Koch[2]

Abstract

■ Most theories of visual processing assume that a target will "pop out" from an array of distractors ("parallel" visual search, e.g., color or orientation discrimination) if targets and distractors can be discriminated without attention. When the discrimination requires attention (e.g., rotated L vs. T or red–green vs. green–red bisected disks), "serial" examination is needed in visual search. Attentional requirements are also frequently assessed by measuring interference from a concurrently performed attentionally demanding task. It is commonly believed that attention acts equivalently in dual-task and visual search paradigms, based on the implicit assumption that visual attentional requirements can be defined along a single dimension. Here we show that there is no such equivalence: We report on targets that do not trigger pop-out, even though they can be discriminated from distractors with attention occupied elsewhere (natural scenes, color-orientation conjunctions); conversely, we show that certain targets that pop out among distractors need undivided attention to be effectively discriminated from distractors when presented in isolation (rotated L vs. +, depth-rotated cubes). In other words, visual search and dual-task performance reveal attentional resources along two independent dimensions. We suggest an interpretation of these results in terms of neuronal selectivities and receptive field size effects. ■

INTRODUCTION

A commonly accepted view of visual processing, largely influenced by the 20-year-old "feature integration theory" of attention (Treisman & Gelade, 1980), holds that our perception of the visual world relies on the parallel extraction of a limited set of preattentive features, followed by the serial integration of these features into a coherent percept, a mechanism mediated by visual attention. In visual search experiments, the detection of "basic" or "preattentive" features can thus be performed in parallel, whereas searching for more complex stimuli such as feature conjunctions typically requires a serial examination by some form of attentional spotlight (Wolfe, Cave, & Franzel, 1989; Treisman & Gelade, 1980). In addition to the classical visual search paradigm, "dual-task" experiments are often used to probe the attentional requirements of discrimination tasks by measuring interference from a concurrently performed attentionally demanding task (Lee, Koch, & Braun, 1999; Braun & Sagi, 1990; Braun, 1993, 1994; Braun & Julesz, 1998; Sperling & Melchner, 1978; Sperling & Dosher, 1986). Although there is no a priori reason to believe so, visual search and dual-task paradigms are commonly assumed to reveal the same attentional resource. ["Fea-

tures not processed in parallel are also not discriminated outside the attentional focus" (Braun, 1998, p. 346).] This relies on the notion that there exists a dichotomy of visual discrimination tasks, with "parallel/preattentive" discriminations on the one hand, and "serial/attentive" discriminations on the other. More recent theories of attention have acknowledged the need to redefine this dualistic terminology, but continue instead to refer to a "continuum" of visual tasks (Norman & Bobrow, 1975) along a single dimension, now entitled "easy vs. difficult" (Nakayama & Joseph, 1998; Joseph, Chun, & Nakayama, 1997) or "efficient vs. inefficient" (Wolfe, 1998).

Here we report that at least two independent dimensions are necessary to describe the space of visual discrimination tasks: one with respect to visual search performance and another with respect to dual-task performance. We provide examples of targets that do not trigger pop-out even though they can be discriminated from distractors in dual-task situations, and of targets that can be searched for in parallel even though they cannot be discriminated from distractors when attention is occupied by a concurrent task. Since it is clear that both visual search and dual-task paradigms are pertinent to the study of attention, we conclude that there must exist at least two distinct visual attentional resources. While the idea that there might exist multiple attentional resources has been considered before (Kahneman, 1973; Treisman, 1969), it has not been widely accepted in visual neuroscience. For clarity, in what

---

[1]CNRS, [2]California Institute of Technology
*These two authors contributed equally to the work reported here.

follows we will refer to the specific attentional resource revealed by visual search as "attention$_{vs}$," and to that revealed by dual tasks as "attention$_{dt}$," with "preattentive$_{vs}$" (or, more intuitively, "parallel") and "preattentive$_{dt}$" denoting visual processes that are not limited by the resource in question.

RESULTS

Visual Search for Some Preattentive$_{dt}$ Features/ Objects Is Serial

Animal versus Nonanimal

It has recently been demonstrated that the discrimination of natural scenes containing one or more animals—insects, mammals, birds, and so on—from scenes that contain no animal can be carried out concurrently with a second attentionally demanding task (visual search among five Ls and/or Ts at the center of the screen) without a significant cost (Li, VanRullen, Koch, & Perona, 2002; see also Figures 2 and 3). Here we show that although this categorization is not limited by attention$_{dt}$, it does not result in parallel performance in a visual search paradigm (i.e., it is in fact limited with respect to attention$_{vs}$).

Eight naïve subjects were asked to detect the presence of a scene containing an animal in an array of up to 16 independent natural scenes. Half of the trials contained a single target scene, whereas the other half contained only distractor scenes. For comparison, the same subjects were also asked in separate blocks to search for the letter L (randomly rotated) among a number of signs + (randomly rotated) or among an array of letters T (randomly rotated). All three symbols (L, +, T) were synthesized from the same two bars perpendicular to each other. The former task is a well-known example of pop-out, whereas the latter is known to require serial search (e.g., Bergen & Julesz, 1983). The subjects were presented with randomly alternating blocks of 96 trials of these three different tasks. In all three tasks, the search array (which could contain 1, 2, 4, 6, 8, 12, or 16 stimuli) was arranged at random locations on a 4 × 4 lattice that subtended roughly 15 × 10 degrees of visual angle. The symbols were about 1.5° in size, and each natural scene subtended roughly 3.5° × 2.5°.

In one experimental condition the search was "response-terminated"; that is, the search array remained visible until the subjects responded by pressing one of two possible keys (one for target-present, the other for target-absent trials). The results presented in Figure 1A show that the search for a natural scene among other scenes is clearly "serial," with search slopes comparable to that of the "L versus T" discrimination: Search slopes were 5 msec/item for the "L versus +" task (target-present trials: 3.4 msec/item; target-absent trials: 6.3 msec/item), 42.5 msec/item in the case of the L versus T task (target-present: 26.2 msec/item; target-absent:

58.1 msec/item), and 40 msec/item for the "animal versus nonanimal" task (target-present: 15.8 msec/item; target-absent: 62.8 msec/item).

We considered the possibility that parallel processing could have nevertheless occurred for natural scenes, but because of the task design, our subjects could have been tempted to wait before responding to make their decisions more reliable when the number of distractors was increased (and thus the signal-to-noise ratio was decreased). By masking all search arrays after 200 msec, we ensured that there would be no advantage of adopting this kind of strategy. Eight other naïve subjects performed this version of the experiment. In this case, the motor response was given in a go/no-go fashion by releasing the mouse button as soon as a target was detected. As expected, the search slopes obtained for all three different tasks were flat (i.e., between 0 and 2 msec/item), and thus the relevant variable was the subjects' percentage correct. The results illustrated in Figure 1B show that here again, the search for a natural scene among others was carried out with a performance comparable to the serial search task (L vs. T), and much different from the parallel one (L vs. +).

Color-Orientation Conjunctions

In fact, a compatible conclusion can be inferred from existing experimental results. It has long been known that searching for a target defined by the conjunction of some simple features such as color and orientation results in serial search slopes (e.g., Treisman & Gelade, 1980). Note that this result has been challenged by more recent studies in which this type of conjunction search (e.g., orientation and color) could be as efficient as any parallel search (e.g., Wolfe et al., 1989). However, this discrepancy in results has not been explained (Wolfe, 1998), and one can reasonably assume that at least under certain conditions color-orientation conjunction search can lead to serial performance. Yet, a recent dual-task experiment by Braun and Julesz (1998) indicated that this type of discrimination can be performed with divided attention. In these experiments, subjects were required to identify peripherally flashed "objects" defined by a conjunction of color and orientation (e.g., a green vertical bar was identified as "zucchini," a red vertical bar as "carrot," a red horizontal bar as "poppies," etc.; each of these nicknames corresponded to a particular keyboard response) while simultaneously performing the attentionally demanding task mentioned previously (visual search among 5 Ts and/or Ls). Whereas one of these stimuli among a combination of the other ones does no always pop out (e.g., Treisman & Gelade, 1980; but see Wolfe et al., 1989), it was perfectly possible for subjects to recognize these stimuli in isolation, even when attention$_{dt}$ was tied down at fixation. The fact that this result chal-

lenges the idea of an absolute equivalence between attention$_{dt}$ and attention$_{vs}$ was not explicitly mentioned by the authors.

Supportive evidence can also be derived from experiments involving the processing of human faces. Numerous visual search studies have demonstrated that despite its obvious biological relevance, facial information (e.g., identity, expression) is not registered in parallel (Brown, Huey, & Findlay, 1997; Purcell, Stewart, & Skov, 1996; Nothdurft, 1993). Yet a recent dual-task experiment indicates that facial information (in this case, face gender) is available to our visual systems even when attention$_{dt}$ is not (Reddy, Wilken, Fried, & Koch, 2002).

In summary, it appears that the absence of requirement with respect to attention$_{dt}$ does not necessarily translate into a parallel visual search. This finding seriously challenges the idea of an equivalence between visual attentional resources in different paradigms. We now proceed to dissociate this equivalence by presenting evidence for parallel features that cannot be discriminated without attention$_{dt}$.

Some Parallel Features/Objects Cannot Be Discriminated without Attention$_{dt}$

Rotated L versus Rotated +

The search for the rotated letter L among many rotated + symbols and the search for the rotated letter L among many rotated Ts are, respectively, two classical examples of parallel and serial visual search (Bergen & Julesz, 1983; see also Figure 1). The latter task is also well known to result in strong interference when performed simultaneously with another attentionally demanding task (Braun & Julesz, 1998). Here we investigate whether the former task (rotated L vs. +) can be performed under similar dual-task conditions.

Five subjects well trained in dual-task experiments (including two of the authors) were presented with rotated Ls or +s briefly flashed in the periphery of the
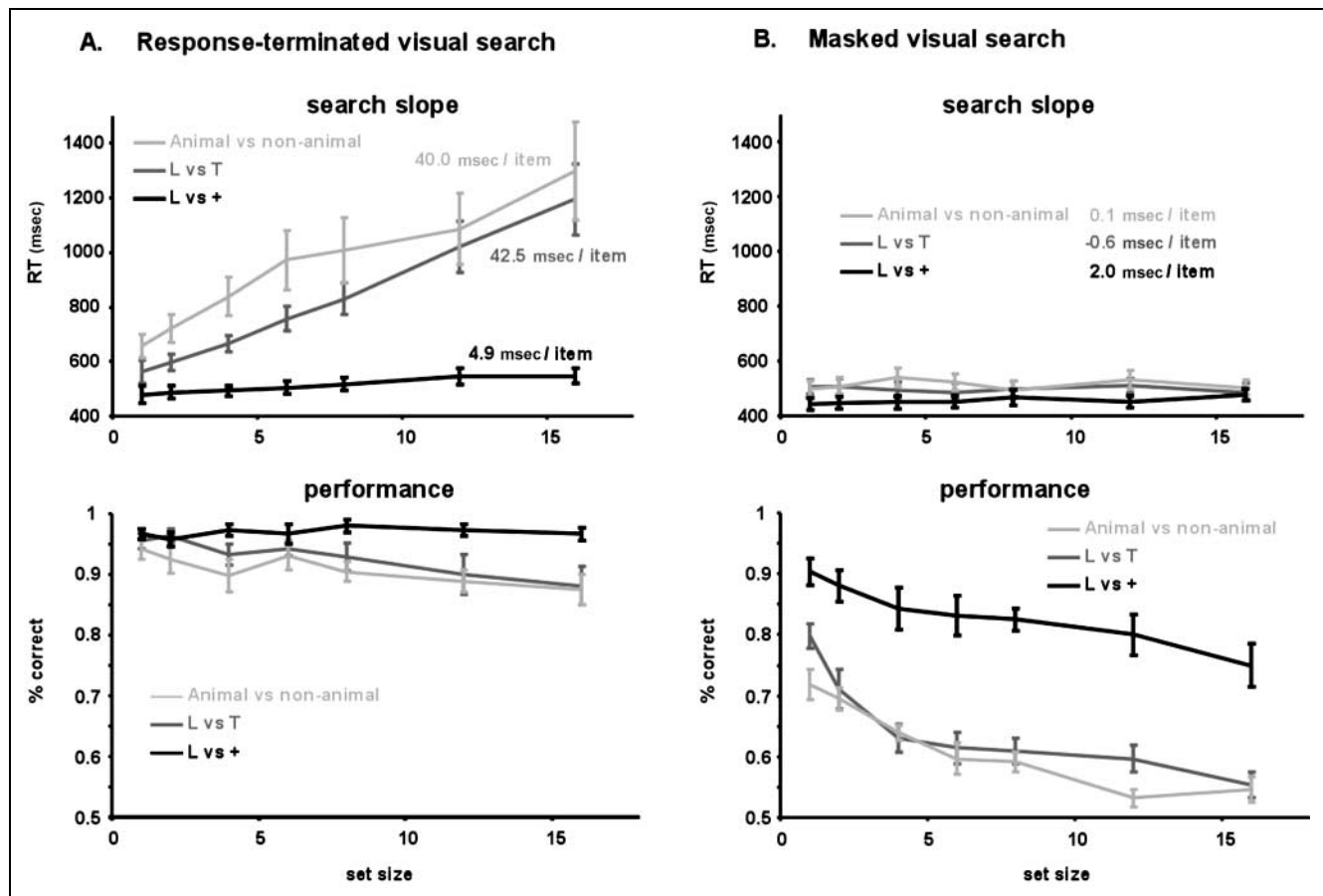

Figure 1. Animals do not pop out. (A) In response-terminated visual search, finding a picture containing an animal among up to 16 other pictures yields search slopes and performance curves typical of serial visual search (40 msec/item). For comparison, performance of the same eight naïve subjects on well-known examples of parallel (rotated L vs. +) and serial (rotated L vs. T) search is also presented. (B) We considered the possibility that subjects might be using longer times to make their decisions not because their visual system actually needed it, but simply because such time was available: Due to a lack of confidence in the animal task, our subjects might have been tempted to wait more when the number of stimuli in the array was increased. When search arrays were masked after 200 msec, subjects (eight other naïve participants) could not adopt a strategy of waiting and accumulating information before making a perceptual judgment. As a result, all search slopes were flat in this case. However, here again the percentage correct in the animal task exhibited the same trend as the one for serial visual search.

visual field and followed by a mask (the stimulus onset asynchrony varied among subjects, between 80 and 120 msec). Five letters (Ts or Ls) were presented at the same time at the center of the screen. In some experimental blocks, the subjects were required to respond to the peripheral stimulus. In other blocks, they had to indicate whether the five central letters were identical or not. In a certain number of blocks, they were asked to perform both tasks simultaneously while fixating at the center of the screen. For comparison, we also asked our subjects to perform a peripheral natural scene categorization task (animal vs. nonanimal) under the same conditions: This task is known to result in only minimal dual-task interference (Li et al., 2002). Note that only two of the five subjects (LR and RV) had previously practiced dual task with this animal versus nonanimal peripheral task. Thus, a systematic difference between this task and the rotated L versus + task cannot be explained simply by the effects of practice.

Results are presented in Figure 2. For each subject there was a dramatic decrease (t test, $p < .0005$) in performance of the peripheral task (rotated L vs. +) between the single- and dual-task conditions. There was no such decrease ($p > .05$) in the peripheral natural scene categorization task (except for subject TJ, $p = .02$). Thus, although the discrimination of Ls among +s is largely accepted as an example of parallel task (see, e.g., Figure 1), the discrimination of the letter L or the + sign cannot be performed when attention$_{dt}$ is not available.

Depth-Rotated Cubes

Perhaps one of the most intriguing examples of parallel search is the search for a target cube among distractor cubes that differ only by their orientation in three-dimensional space: Despite the seemingly complex or ''high-level'' nature of this task, reaction times were found to increase only minimally with set size (Sun & Perona, 1996; Enns & Rensink, 1990). Here we report that such target and distractor cubes cannot be discriminated when attention$_{dt}$ is unavailable. The experimental paradigm was identical to the one described in the previous section. In short, the same five subjects were required to discriminate between two possible depth-rotated cubes under dual-task conditions. They performed this task in alternation with an animal versus nonanimal task. The cube stimuli were similar to those described by Sun and Perona (1996). Results are presented in Figure 3, and show a strong ($p < .005$) decrease of performance of the depth-rotated cubes discrimination task between the single- and dual-task conditions.

DISCUSSION

The evidence reported here challenges a common assumption about visual attention: that preattentive$_{dt}$ and parallel processing are two equivalent ways to refer to the same subset of visual discrimination tasks. This is obviously wrong, because some tasks can be preattentive$_{dt}$ but not parallel, and vice versa. In a more constructive fashion, these results call for a refinement of the accepted terminology of visual attention. We propose one interpretation of these results based on the neuronal processes likely to be involved in the different discrimination tasks reviewed, in particular, neuronal selectivities and receptive field sizes. This need not be the only valid explanation, but it appears to us to be the most sensible in light of our results.

Refining the Terminology of Visual Attention

What determines whether a task requires attention$_{dt}$?

Our results do not directly challenge the notion of preattentive features (of the type proposed by Treisman & Gelade, 1980, and Wolfe et al., 1989). But the present results (together with those of Li et al., 2002) show that such features cannot be labeled with a certain level of complexity (i.e., are not limited to early cortical representations), and that one should instead refer to a hierarchy of preattentive$_{dt}$ features (this idea was first introduced by Treisman herself, in 1988). In neural terms, such preattentive features appear to be represented at various stages of the ventral pathway, from V1 to IT. To account for these results, we propose that a feature, stimulus or object category is preattentive$_{dt}$ if there exists a neuronal population selective to this feature, stimulus, or object category, independent of the cortical area involved (at least within the ventral visual stream). Indeed, neurons in V1 are well known for their selectivity to orientation (in columns) or color (in blobs; Zeki, 1983), both feature dimensions being clearly preattentive$_{dt}$ (e.g., Braun & Julesz, 1998). Similarly, neurons selective for simple conjunctions of color and orientation (such as carrots, zucchini, poppies, etc.) can most likely be found in areas beyond V4 (Ghose & Ts'o, 1997; Gallant, Braun, & Van Essen, 1993; Desimone, Schein, Moran, & Ungerleider, 1985; Zeki, 1983) or PIT (Tanaka, 1993, 1996), whereas neurons in the more anterior parts of the inferotemporal cortex in monkeys or in the medial temporal lobe in humans can be highly selective for particular categories of objects such as animals or faces (Kreiman, Koch, & Fried, 2000; Desimone, Albright, Gross, & Bruce, 1984; Perrett, Rolls, & Caan, 1982; Gross, Rocha-Miranda, & Bender, 1972). These selectivities could ensure that the corresponding discrimination tasks (e.g., animal vs. nonanimal, carrot vs. zucchini) can be performed without attention$_{dt}$.

Of course, we are not suggesting that in the case of attentive$_{dt}$ tasks, there exists no neuronal population that can potentially discriminate between the target and distractors (for in that case, how could subjects ever perform these tasks?). What we mean is precisely that for such tasks, visual attention$_{dt}$ is necessary to construct the relevant neuronal selectivities. According to our

Figure 2. Rotated L and + cannot be discriminated in the near absence of attention$_{dt}$ (gray diamonds). For comparison, the preattentive$_{dt}$ animal versus nonanimal natural scene categorization task was performed by the same subjects in separate blocks during the same experimental sessions (black circles). (A) On each trial, five letters (Ts or Ls) were flashed at the center of the screen, and a stimulus appeared at a random location in the periphery after 53 msec. Both central and peripheral stimuli were masked after presentation, and the SOA was determined individually for each subject (varying between 133 and 200 msec in the central task, and between 80 and 120 msec in the peripheral task). In some experimental blocks (gray diamonds), the peripheral stimulus was either the letter L or the + sign (randomly rotated), while in other blocks (black circles), the peripheral stimulus was a natural scene containing one or more animals on half the trials. On different blocks, the subjects were instructed to perform either the central (Are the five letters identical or different?) or the peripheral (Is it an L or a +? Is there an animal or not?) tasks in isolation, or to perform both tasks simultaneously. The results are presented in the form of attention-operating characteristics (AOC), with central performance plotted on the horizontal axis and peripheral performance on the vertical axis. By convention, in the single-task condition the performance level of the "to-be-ignored" task is set to chance level (50%). (B) Individual results show that although most subjects can perform the animal versus nonanimal categorization task simultaneously with the central letter discrimination task without a significant cost ($p > .05$; except for subject TJ, $p = .02$), their performance of the peripheral rotated L versus + discrimination task suffers considerably ($p < .0005$) in the ual-task condition. (C) These results are summarized in a normalized AOC plot. Each black circle or gray diamond corresponds to a particular subject's performance in dual-task, normalized to the corresponding level of performance in the single-task condition (a linear transformation brings single-task performance to 100% while leaving chance level at 50%). The difference between the levels of performance achieved with little or no attention$_{dt}$ in animal versus nonanimal task and the rotated L versus + task is particularly apparent here.
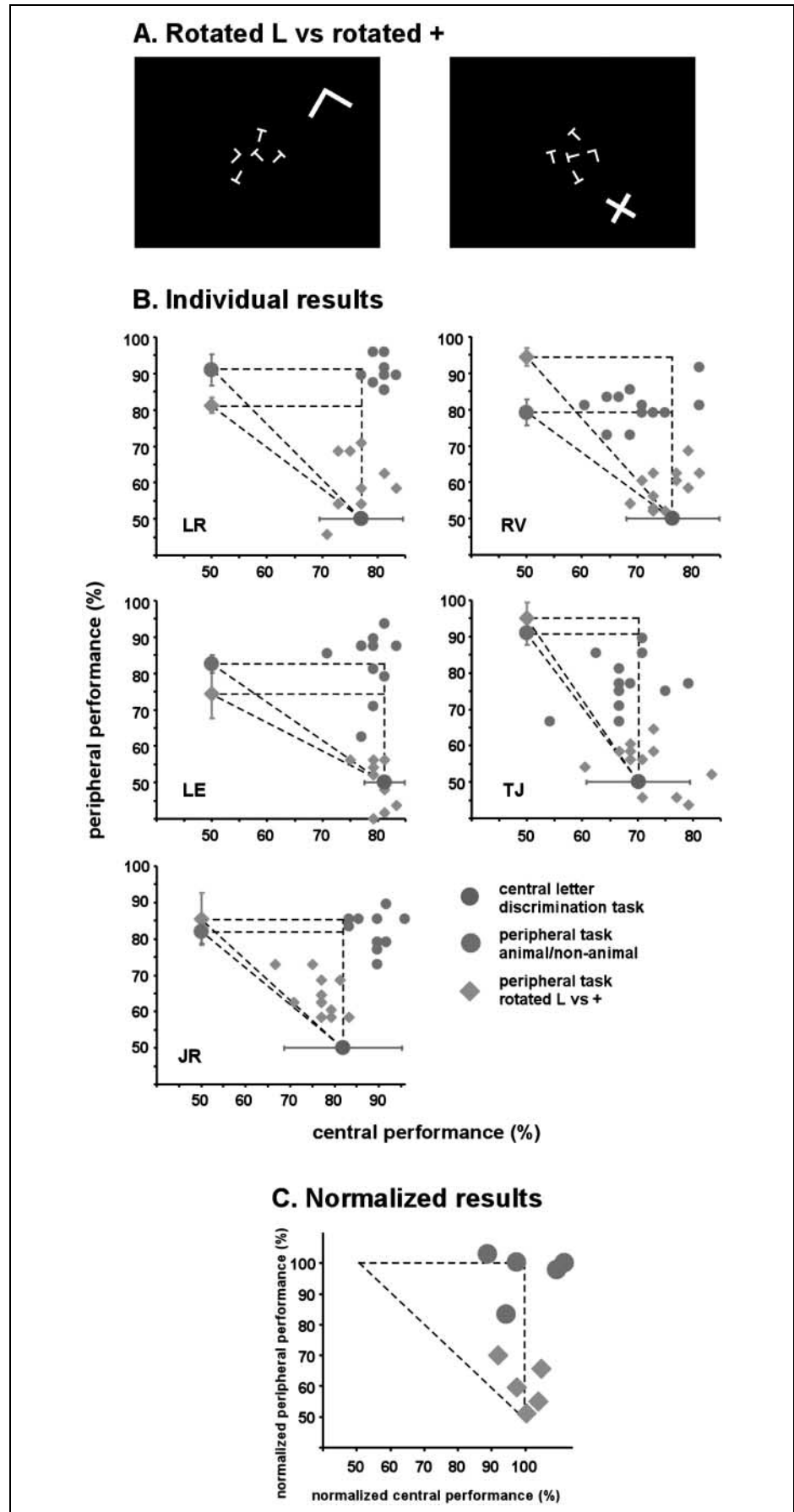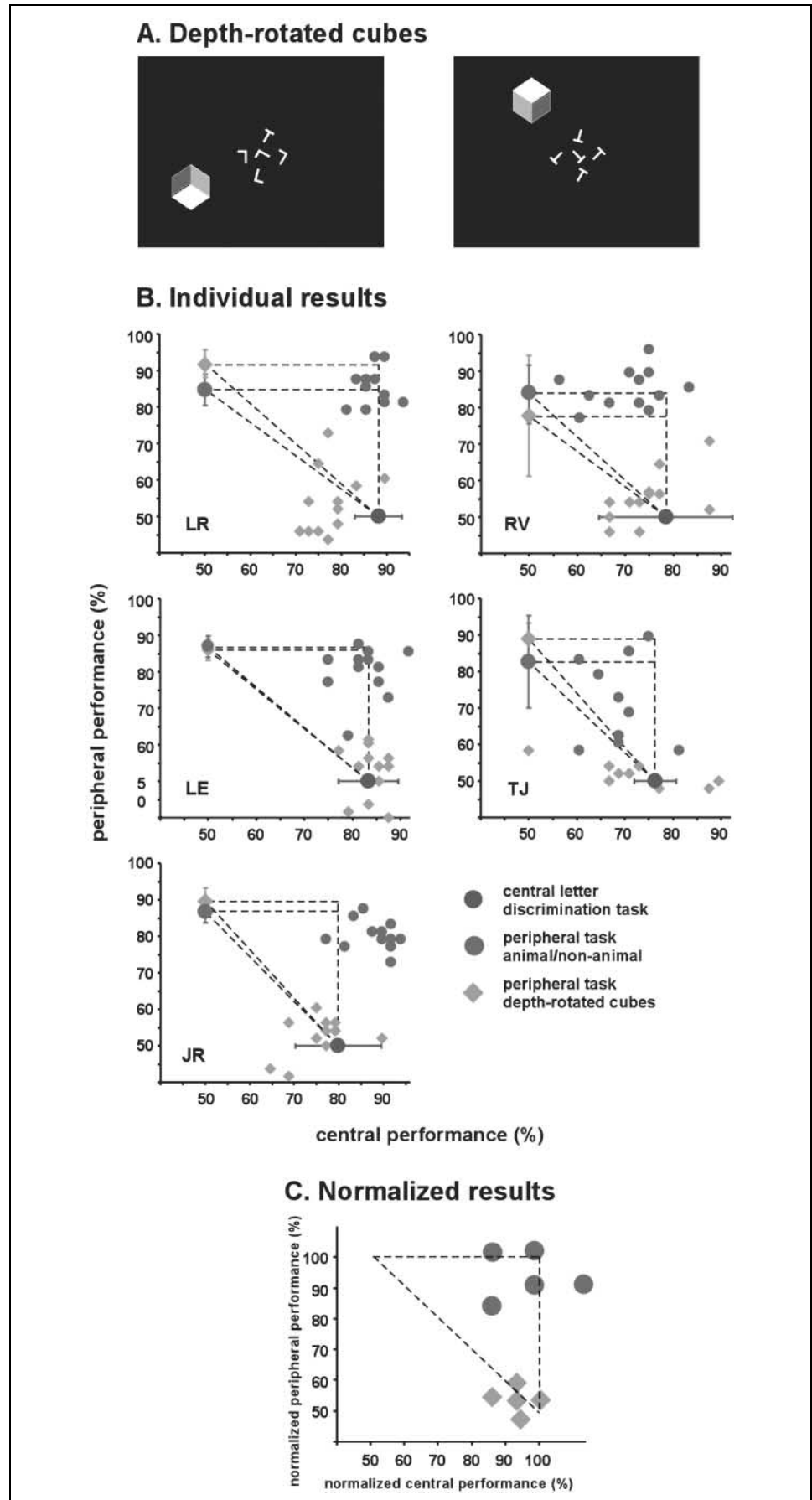
Figure 3. (A) Depth-rotated cubes cannot be discriminated in the near absence of attention$_{dt}$ (gray diamonds). As previously, the preattentive$_{dt}$ animal versus nonanimal natural scene categorization task was performed (for comparison) by the same subjects during the same experimental sessions (black circles). The SOA for the central task and peripheral animal versus nonanimal task was the same as in the experiment depicted in Figure 2, while the SOA for the peripheral depth-rotated cubes task varied between 120 and 200 msec. Notations as in Figure 2. (B) Here again, individual results show that whereas our subjects can detect the presence of an animal in a natural scene while simultaneously performing the central task without a significant cost ($p > .05$), their performance of the peripheral depth-rotated cubes discrimination task is greatly impaired ($p < .005$) under the same conditions. (C) Normalized AOC plot.

results and the previous literature (Lee et al., 1999; Braun, 1998), this would be the case when discriminating rotated letters (L vs. T or +), depth-rotated cubes, or bisected two-color disks. It is likely that for such tasks, attention$_{dt}$ is required to mediate a more detailed analysis, possibly involving mental rotations or binding.

Other authors had previously considered relating preattentive processes to the "feed-forward visual system" (Lamme & Roelfsema, 2000), but could not pursue this idea because of the consistent failure of demonstrating parallel search for simple feature conjunctions or objects. In other words, some features (e.g., faces) corresponding to feed-forward neuronal selectivities did not appear to yield preattentive$_{vs}$ processing. Here we propose that such an equivalence can in fact be estab-
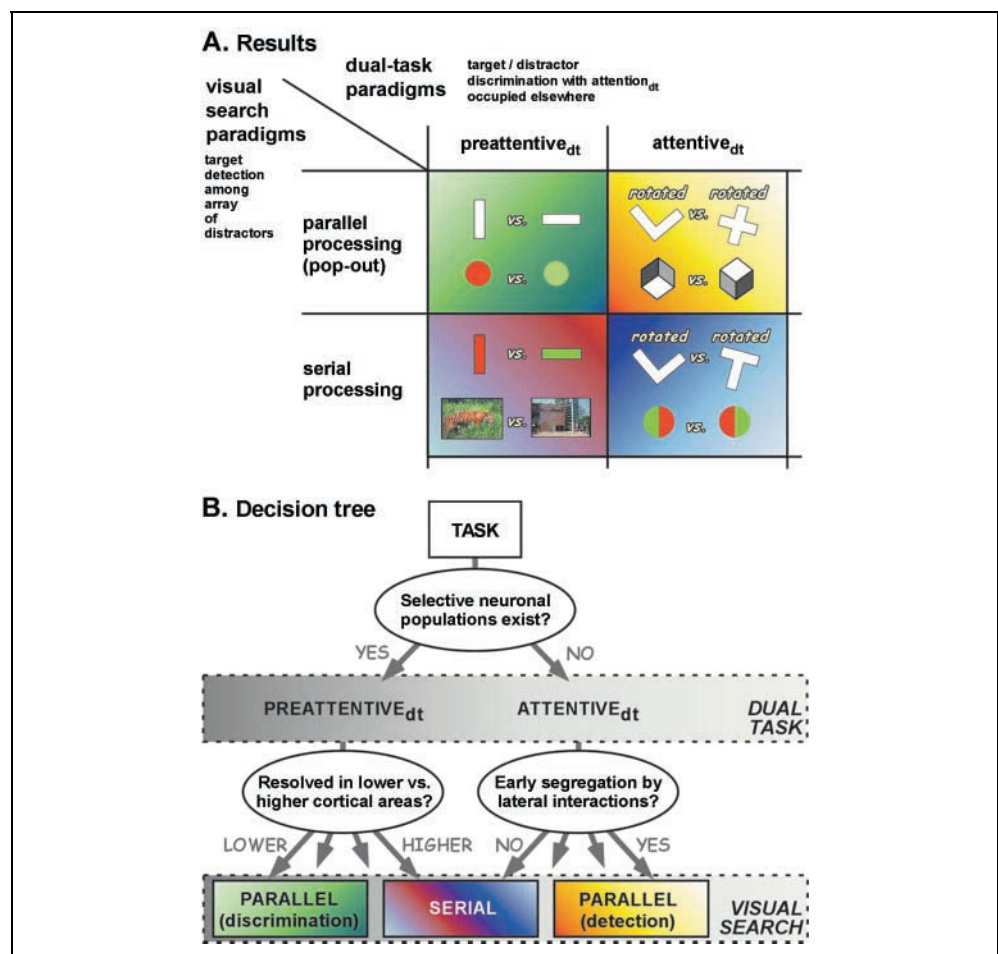
lished when only one attentional resource, attention$_{dt}$, is considered.

In theory, preattentive$_{dt}$ implies parallel, but in practice, the latter is constrained by the size of receptive fields

In the terminology of Treisman and Gelade (1980) or Wolfe et al. (1989), preattentive features can be processed in "parallel." We believe this to be the case but (because of the preceding point) with particular constraints. We note that in the examples presented in this paper (see summary in Figure 4), the "preattentive$_{dt}$" tasks that result in parallel visual search seem to rely on neuronal selectivities present in early visual areas

Figure 4. Summary of results and hypothesis. (A) Two independent dimensions are needed to account for the variety of visual discrimination tasks: one with respect to visual search performance (the "parallel versus serial" dimension), the other with respect to dual-task performance (the "preattentive$_{dt}$ vs. attentive$_{dt}$" dimension). In addition to well-known tasks that can be defined as "both parallel and preattentive$_{dt}$" (e.g., orientation or color discrimination; Braun & Julesz, 1998; Treisman and Gelade, 1980) or "both serial and attentive$_{dt}$" (e.g., rotated L vs. T, two-color bisected disks; Lee et al., 1999; Bergen & Julesz, 1983), we presented experimental evidence for "serial and preattentive$_{dt}$" visual tasks (animal vs. nonanimal natural scene categorization, or color-orientation conjunction; Li et al., 2002; Braun & Julesz, 1998; Treisman & Gelade, 1980) as well as "parallel and attentive$_{dt}$" tasks (rotated L vs. +, depth-rotated cubes; Sun and Perona, 1996; Enns and Rensink, 1990; Bergen & Julesz, 1983). (B) Decision tree accounting for the different

classes of discrimination tasks, as revealed by dual tasks and visual search. We propose that a task is preattentive$_{dt}$ if there exists neuronal populations that respond selectively to targets and distractors, independent of the cortical area involved. Without such selective neuronal populations, attention$_{dt}$ will be necessary to dynamically construct the required selectivities, for example, by mediating further operations such as binding or mental rotations. Pop out (parallel processing) relies in general on neural mechanisms taking place in early visual areas (V1–V2), where small receptive fields (RFs) and lateral or feedback interactions outside the classical receptive field (e.g., grouping, segmentation) can work to isolate the target from the distractors. If the corresponding discrimination task is preattentive$_{dt}$, these early mechanisms can result in the parallel *discrimination* of targets and distractors (e.g., orientation, color); otherwise, the presence of an odd item can still be *detected* in parallel (pop-out effect), but its identity will not be registered in the absence of attention$_{dt}$ (rotated L vs. +, depth-rotated cubes). "Intermediate" arrows at the last level illustrate the fact that visual search performance appears to lie on a continuum (Nakayama & Joseph, 1998; Wolfe 1998; Norman & Bobrow, 1975).

(e.g., orientation, color), while those that result in serial visual search probably rely on higher level neuronal selectivities (color-orientation conjunctions, animals, faces). These differences in neuronal selectivities are usually accompanied by differences in the size of the neuronal receptive fields (Desimone, Moran, & Spitzer, 1988). Thus we propose that the extent to which our preattentive$_{dt}$ features can be discriminated in parallel is an inverse function of the receptive field size of the neurons that represent this feature. At higher levels of the ventral hierarchy, only very few "features" can be processed in parallel, and the corresponding stimuli must be well enough separated to avoid having a target and a distractor falling within a single receptive field. This could explain why a recent study by Rousselet, Fabre-Thorpe, and Thorpe (2002) found that two natural scenes can be processed in parallel (in an animal vs. nonanimal categorization task) when they are presented to the left and right of fixation, and thus activate very distinct neuronal populations (most probably in different hemispheres). With receptive fields covering the entire visual field, there could be no parallel processing, even when the target is defined as a preattentive$_{dt}$ feature. This idea is still speculation at this time, but it can be experimentally tested because it predicts that performance in visual search for preattentive$_{dt}$ features should increase as display density decreases—a property that has been verified for simple feature conjunctions (Berger & McLeod, 1996; Cohen & Ivry, 1991), but could be more systematically investigated. Note that this idea could also explain why certain studies find that color-orientation conjunction search is serial (e.g., Treisman & Gelade, 1980), and other studies find the same conjunction search to be parallel (e.g., Wolfe et al., 1989): Interstimulus distance and receptive field clutter might account, at least in part, for this discrepancy.

Parallel does not imply preattentive$_{dt}$

We have shown that pop-out can also occur for features that are not discriminated without attention$_{dt}$ (such as an L among +s, or depth-rotated cubes). What then underlies this parallel search, if we assume that attention$_{dt}$ is a limited resource that cannot be deployed simultaneously across the entire search array? In this case, we propose that parallel search might rely on center-surround and lateral interaction mechanisms occurring at early stages of the cortical hierarchy (Gilbert, Das, Ito, Kapadia, & Westheimer, 1996; Gilbert, 1992; Ts'o, Gilbert, & Wiesel, 1986; Allman, Miezin, & McGuinness, 1985). Examples of such possible mechanisms include distractor grouping (Duncan & Humphreys, 1989), surface integration (He & Nakayama, 1992; Nakayama & Shimojo, 1992) and texture segmentation (Julesz, 1975, 1981, 1986): each of these mechanisms has already been demonstrated to facilitate "efficient"

visual search. The common factor among these integration processes is that they can reveal the location of a "salient" item (e.g., an item that "sticks out" of the texture or surface defined by surrounding elements) without necessarily having direct access to its identity. In other words, the presence of an odd element, rather than its identity, could mediate this pop-out effect. In fact, this type of pop-out display (i.e., the entire search array, with the target "sticking out" of the surrounding items) might itself be thought of as a preattentive feature, at a rather coarse scale of processing (Nakayama & Joseph, 1998; see, e.g., Braun, 1993; Braun & Julesz, 1998). The distinction between this type of parallel *detection* of an odd item and the parallel *discrimination* of targets and distractors described previously (Point 2) has already been suggested by Wolfe (1992; see also Duncan & Humphreys, 1989; Sagi & Julesz, 1985; Beck, 1966).

In practice, dual-task performance can reveal whether a visual feature or property is preattentive$_{dt}$ (Point 1), whereas visual search performance can reflect both parallel *discrimination* (which requires that a preattentive$_{dt}$ feature be represented in early enough areas with small receptive fields; Point 2) and low-level grouping and segmentation mechanisms that work to isolate the target but are blind to its identity (Point 3). It is no surprise therefore that there appears to be a continuum of visual search performance (Nakayama & Joseph, 1998; Wolfe, 1998; Duncan & Humphreys, 1989), reflecting the relative participation of each mechanism, and the fact that parallel discrimination could vary gradually with the cortical level of representation of the preattentive$_{dt}$ feature (and thus the receptive field size; Treisman & Gormican, 1988).

The theory sketched here in coarse lines leaves aside, for now, a number of observations. For instance, our account of preattentive orientation discrimination concentrates on first-order, luminance-defined orientation. There is ample evidence that other, second-order mechanisms based on texture or color cues can support efficient orientation processing (e.g., Bravo & Blake, 1990; Cavanagh, Arguin, & Treisman, 1990), and that categorical properties such as "steepness" versus "shallowness" can also facilitate visual search (Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992). A full account of attention and visual search would have to consider these and other mechanisms.

Our results will hopefully help to make sense of the large body of (often conflicting) experimental data on visual attention. Specifically, our interpretation suggests that the role of visual attention can be twofold: (1) to dynamically generate neuronal selectivities that are not explicitly implemented in the visual system at the level of single neurons (a role compatible with the "feature-integration theory" of attention and related proposals; Wolfe et al., 1989; Treisman & Gelade, 1980) and/or (2) to resolve spatial ambiguities that arise when multiple stimuli fall into the same receptive field (a role similar

to that proposed in the "biased-competition framework"; Desimone & Duncan, 1995; Moran & Desimone, 1985; see also Mozer & Sitton, 1998; Motter, 1993). The present results illustrate the conditions under which each of these two attentional mechanisms can be revealed.

METHODS

Visual Search Experiments

The subjects were seated in a dark room, 120 cm from a computer screen piloted from an Apple G4 computer. They performed three different search tasks in alternating blocks of 96 trials. In one task, they had to search for a natural scene containing one or more animals (mammals as well as insects, reptiles, birds, or fish) among a number of scenes containing no animal (e.g., landscapes with trees, city scenes with buildings, vehicles, indoor scenes, etc.); in the other two tasks, they were required to search for the letter L among a number of Ts or a number of + signs. All three of these symbols were synthesized from the same two bars perpendicular to each other, and randomly rotated. The symbols subtended roughly 1.5°, while each natural scene subtended approximately 3.5° × 2.5°. In all three tasks, the search array could contain 1, 2, 4, 6, 8, 12, or 16 stimuli. In half of the trials the search array contained one target. The stimuli were randomly arranged on a 4 × 4 lattice that subtended 15° × 10°. Their exact position with respect to the lattice was slightly shifted randomly in the horizontal and vertical directions to avoid the effects of stimulus alignment.

In one condition, 8 naïve subjects were required to perform the three search tasks in response-terminated conditions: They were asked to press one key when the target was present, and another when the target was absent, the search array remaining visible until the response.

In another condition, 8 other naïve subjects were asked to perform the three search tasks in a go/no-go fashion: They held down the mouse button to run the sequence of trials; each stimulus in the array was masked 200 msec after presentation; the subjects were required to release the mouse button as fast as possible when they detected a target, and to keep pressing it otherwise. For the three symbols (L, +, T), the mask was another symbol made of a combination of the target and distractor stimuli, and rotated by the same angle as the masked symbol. For the natural scenes, eight different masks were used, each one a mixture of white noise at different spatial frequencies on which a naturalistic colored texture was superimposed.

Dual-Task Experiments

The subjects were seated in a dark room, 120 cm from a computer screen piloted from a SGI O2 workstation. All 5 subjects, including two of the authors, were well trained in dual-task experiments. Each trial was organized as follows. The subjects were asked to fixate on a central cross. At time $t = 0$, the cross disappeared and was replaced by five randomly rotated letters placed randomly on a 3 × 3 lattice occupying roughly 2° at the center of the screen; the letters could be 5 Ls, 5 Ts, 4 Ls and 1 T, or 4 Ts and 1 L. After 53 msec, a stimulus was flashed (for 27 msec) in the periphery, at a random location centered on an imaginary rectangle subtending 10° × 8°; thus the eccentricity of the peripheral stimulus was on average between 4° and 5°. In the natural scene categorization task, the peripheral stimulus was a natural scene (subtending roughly 5° × 3°) that contained one or more animals on half of the trials. In the rotated L versus + discrimination task, the peripheral stimulus was either the letter L or the + sign (with equal probability). Both symbols were approximately 1° in size, and randomly rotated on each trial. In the depth-rotated cubes task, the peripheral stimulus was a shaded cube (measuring approximately 1.5°) of two possible orientations in depth (which were equiprobable), each with a different illumination angle (see Figure 3A). Each peripheral stimulus was masked after a stimulus onset asynchrony (SOA) that was determined individually for each subject to ensure a performance level of approximately 80% on the single peripheral task. For the natural scene categorization task, eight different masks were used, each one a mixture of white noise at different spatial frequencies on which a naturalistic colored texture was superimposed. For the rotated L versus + task, the mask was a combination of the two symbols, rotated with the same angle as the masked stimulus. For the depth-rotated cubes task, the mask was a collage of parts of the two stimulus cubes. The five central letters were also masked after a "central SOA" that was determined individually for each subject to ensure a performance level of the central task of about 80%. In the "central task" condition, the subjects were required to focus at the center of the screen and ignore the peripheral stimuli; they were asked to press one key if the central five letters were identical, and another if one letter was different from the others. In the "peripheral task" condition, subjects were asked to fixate at the center of the screen and discriminate the peripheral stimuli. Depending on their previous dual-task experience, three of the five subjects responded to the target stimulus by releasing the mouse button (no response was required for the distractor stimulus), while the other two (LE and TJ) responded using one key on target trials and another on distractor trials. (Note that the results obtained with these two response modes were compatible.) For the natural scene categorization task, the targets were scenes containing one or more animals. For the rotated L versus + task, the target was the letter L. For the depth-rotated cubes task, the target was the cube facing down and illuminated from below. Under the dual-task

condition, the subjects were asked to fixate at the center and perform both the central and peripheral tasks simultaneously. They were instructed to keep their performance of the central task at comparable levels in the single- and dual-task conditions. The experiment was performed in three 1-h daily sessions, in which subjects alternated between the single central, single peripheral, and dual-task conditions; in all three conditions, the rotated L versus + and the depth-rotated cubes tasks were always performed in alternation with the animal versus nonanimal task.

Performance levels in blocks of the single- and dual-task conditions were compared using a *t* test. For the animal versus nonanimal task, the comparison between single- and dual-task conditions was made over all blocks, whether they were presented in alternation with the rotated L versus + or the depth-rotated cubes blocks.

Acknowledgments

REFERENCES

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local–global comparisons in visual neurons. *Annual Review of Neuroscience, 8,* 407–430.

Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science, 154,* 538–540.

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303,* 696–698.

Berger, R. C., & McLeod, P. (1996). Display density influences visual search for conjunctions of movement and orientation. *Journal of Experimental Psychology. Human Perception and Performance, 22,* 114–121.

Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a "texton." *Spatial Vision, 7,* 311–322.

Braun, J. (1994). Visual search among items of different salience: Removal of visual attention mimics a lesion in extrastriate area V4. *Journal of Neuroscience, 14,* 554–567.

Braun, J. (1998). Divided attention: Narrowing the gap between brain and behavior. In R. Parasuraman (Ed.), *The attentive brain* (pp. 327–351). Cambridge: MIT Press.

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

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

Bravo, M., & Blake, R. (1990). Preattentive vision and perceptual groups. *Perception, 19,* 515–522.

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

Cavanagh, P., Arguin, M., & Treisman, A. (1990). Effect of surface medium on visual search for orientation and size features. *Journal of Experimental Psychology. Human Perception and Performance, 16,* 479–491.

Cohen, A., & Ivry, R. B. (1991). Density effects in conjunction search: Evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology. Human Perception and Performance, 17,* 891–901.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4,* 2051–2062.

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

Desimone, R., Moran, J., & Spitzer, H. (1988). Neural mechanisms of attention in extrastriate cortex of monkeys. In M. Arbib & S. Amari (Eds.), *Dynamic interactions in neural networks: Models and data* (pp. 169–182). New York: Springer-Verlag.

Desimone, R., Schein, S. J., Moran, J., & Ungerleider, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Research, 25,* 441–452.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247,* 721–723.

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science, 259,* 100–103.

Ghose, G. M., & Ts'o, D. Y. (1997). Form processing modules in primate area V4. *Journal of Neurophysiology, 77,* 2191–2196.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron, 9,* 1–13.

Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences, U.S.A., 93,* 615–622.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology, 35,* 96–111.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359,* 231–233.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature, 387,* 805–807.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 232,* 34–43.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290,* 91–97.

Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics, 54,* 245–251.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

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

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

Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: Concurrent discrimination of form, color, and motion. *Perception & Psychophysics, 61,* 1241–1255.

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, U.S.A., 99,* 9596–9601.

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

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology, 70,* 909–919.

Mozer, M. C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). London, UK: University College London Press.

Nakayama, K., & Joseph, J. S. (1998). Attention, pattern recognition and pop-out in visual search. In R. Parasuraman (Ed.), *The attentive brain* (pp. 279–298). Cambridge: MIT Press.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257,* 1357–1363.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44–64.

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

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329–342.

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

Reddy, L., Wilken, P., Fried, I., & Koch, C. (2002). *The attentional requirements of face gender discrimination.* Paper presented at the Annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

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

Sagi, D., & Julesz, B. (1985). ''Where'' and ''what'' in vision. *Science, 228,* 1217–1219.

Sperling, G., & Dosher, B. (1986). Strategy and optimization in human information processing. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–65). New York: Wiley.

Sperling, G., & Melchner, M. J. (1978). The attention operating characteristic: Examples from visual search. *Science, 202,* 315–318.

Sun, J., & Perona, P. (1996). Early computation of shape and reflectance in the visual system. *Nature, 379,* 3165–168.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262,* 685–688.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19,* 109–139.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review, 76,* 282–299.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 40,* 201–237.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

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

Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience, 6,* 1160–1170.

Wolfe, J. M. (1992). ''Effortless'' texture segmentation and ''parallel'' visual search are not the same thing. *Vision Research, 32,* 757–763.

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

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology. Human Perception and Performance, 15,* 419–433.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology. Human Perception and Performance, 18,* 34–49.

Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 217,* 449–470.