Binding hardwired versus on-demand feature conjunctions

Rufin VanRullen

Université de Toulouse, CerCo, UPS, and CNRS, UMR5549, Faculté de Médecine de Rangueil, Toulouse, France

Binding denotes the process by which features represented in early stages of the visual system are brought together within a meaningful representation that can support recognition, action, or consciousness. It is often considered that binding requires attentional resources. However, recent results have demonstrated that recognition of natural and familiar objects or scenes—as opposed to more arbitrary, meaningless conjunctions of features—can be done while attention is distracted. Surprisingly, the same discriminations do not lead to efficient “pop-out” visual search. These seemingly opposite conclusions can be reconciled by postulating the existence of two modes of binding in the brain: “Hardwired” binding of frequently encountered, natural objects through a hierarchy of increasingly complex feature detectors, as in classic computational models inspired by neurophysiological results; and “on-demand” binding mediated by attention for more arbitrary or meaningless feature conjunctions, as advocated in classic psychological theories. The former system can function without attention, but is challenged by spatial competition within neuronal receptive fields when multiple objects are presented. Thus, while on-demand binding always requires attention, hardwired binding only does when receptive field competition occurs. New experimental data using multiple stimuli and variable interstimulus distance supports this theoretical dissociation.

Keywords: Attention; Natural scenes; Feed-forward model; Competition; Pre-attentive recognition

When contemporary scientists describe “the binding problem” to less knowledgeable peers, they generally (1) evoke the well-known fact that early stages of visual processing decompose the visual scene into a number of elementary features represented in separate maps (colour, orientation, motion direction, and so on), and then (2) conjure up an example in which, say two cars driving past one another might be easily confounded (does the red paint belong to the eastbound or the westbound car?). Separated features must be “bound” back together at some point.

Please address all correspondence to Rufin Vanrullen, CNRS, UMR5549, Faculté de Médecine de Rangueil, 31062 Toulouse, France. E-mail: rufin.vanrullen@cerco.ups-tlse.fr
Accounts might diverge, however, regarding the nature of the feature binding process. For computational neuroscientists, higher level neurons can “simply” bind elementary features into hardwired\(^1\) combinations, within the limited spatial extent of their receptive fields. Many state-of-the-art models of object recognition are based on such feedforward feature combinations—and the approach works remarkably well (Delorme & Thorpe, 2001; Fukushima & Miyake, 1982; Riesenhuber & Poggio, 1999b; Serre et al., 2005; Serre, Oliva, & Poggio, 2007; VanRullen, Gauthrais, Delorme, & Thorpe, 1998; VanRullen & Thorpe, 2002). Many cognitive scientists, on the other hand, tend to consider that the sum of all arbitrary conjunctions of two or more elementary features must be a large number (the “combinatorial explosion”), too large for each conjunction to merit its own “map”. They are thus led to the unavoidable conclusion that higher level perception must rely not on any “hardwired” binding of features into objects, but on a flexible, “on-demand” strategy. Attention is generally offered as the key to solving the binding problem (Treisman, 1996). This would imply that no object recognition, no scene comprehension or categorization can occur outside of the focus of attention, a classic view that has been supported by much experimental evidence (Chun & Marois, 2002; Nakayama & Joseph, 1998; Treisman & Sato, 1990; Treisman & Souther, 1985; Wolfe, 1998; Wolfe & Bennett, 1997). But upon more careful scrutiny, there are also findings that are difficult to reconcile with this framework. Attentional deployment should take time and effort, yet natural scene or object categorization can be remarkably fast (Kirchner & Thorpe, 2006; Potter & Faulconer, 1975; Potter & Levy, 1969; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001); more recent work even suggests that it could occur with little attentional effort (Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002; Reddy, Reddy, & Koch, 2006; Reddy, Wilken, & Koch, 2004; Rousselet, Fabre-Thorpe, & Thorpe, 2002; VanRullen, Reddy, & Koch, 2004). Is the dogma floundering? Yes, but only partly. I will present evidence that some forms of object recognition, but not all, can happen without attention. The picture will be further complicated by the finding that different attentional paradigms lead to conflicting results. But in the end, I will show that most of the available data can be reconciled simply by abandoning one of the premises of the previous reasoning: By accepting that hardwired binding, as it is implemented in current computational models, can be a viable strategy for at least some types of object recognition (Riesenhuber & Poggio, 1999a). After all, the complex object selectivities found in inferotemporal neurons (e.g., face-selective cells) (Logothetis & Sheinberg, 1996; Perrett, Rolls, & Caan, 1982) are often preserved in

\(^1\) The term “hardwired” does not imply an innate origin. Here, it indicates a pattern of neuronal connectivity that is, in most cases, the result of experience, and which does not alter significantly from one stimulus presentation to the next.
anesthetized animals, when attention and other top-down factors can only play a minor role (Desimone, Albright, Gross, & Bruce, 1984; Gross, Rocha-Miranda, & Bender, 1972; Kobatake & Tanaka, 1994; Tanaka, 1996).

**RECOGNITION WITHOUT ATTENTION**

For decades now, attention has been considered as a necessary condition for visual object recognition. In the debate between early (Bergen & Julesz, 1983; Broadbent, 1958; Hillyard, Vogel, & Luck, 1998; Mangun, 1995; Treisman, 1960; Treisman & Gelade, 1980) and late (Allport, Tipper, & Chmiel, 1985; Deutsch & Deutsch, 1963; Duncan, 1980, 1984; Tipper, 1985) selection theories of attention, early selection has more or less taken the prize. This is in no small part the result of the pioneering research programmes of Anne Treisman and Bela Julesz, using visual search arrays (Bergen & Julesz, 1983; Julesz, 1975; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Paterson, 1984; Treisman & Souther, 1985; Wolfe, 1994, 1998; Wolfe, Cave, & Franzel, 1989). They clearly and repeatedly demonstrated that the visual system can simultaneously register and compare simple features (oriented bars, coloured patches, moving dots) across multiple spatial locations, as if it was an effortless, cost-free problem; but when more complex objects (i.e., combinations of two or more features such as bicoloured items, Ls vs. Ts, colour-orientation conjunctions, etc.) had to be processed, a performance cost was observed for each additional item in the array, as if some limited resource was becoming more and more in demand: Attention. The interpretation was almost evident: The elementary features being represented in separate, retinotopically organized maps (in early cortical areas) could support the simultaneous representation of multiple items; but no such map existed for the more complex objects and feature conjunctions (due to the “combinatorial explosion” problem), and the required representation had to be constructed on demand—a process mediated by attention, which could only occur for one or a few items at a time. In other words, attention served to select what features of the early, parallel representation stages would be bound into higher level representations (the so-called “object files” of Treisman’s theory). But this idea derived from visual search experiments might not have made such an impact if it wasn’t also supported by other experimental techniques. For example in dual tasks (another classic and elegant attentional manipulation in which attentional resources are taken away from a target stimulus by a secondary, challenging task), it was observed that a single elementary feature (horizontal vs. vertical bar; red vs. green patch) could be distinguished as target, but a more complex conjunction (L vs. T, bisected red-green vs. green-red disks) could not (Braun & Julesz, 1998; Braun & Sagi, 1990; Lee, Koch, & Braun, 1999). With such a consistent pattern of results,
most vision scientists agreed that attention must indeed matter for binding features, and thus for object recognition.

What a surprise then when, more recently, a number of studies started to suggest that this outcome could change if natural\(^2\) scenes and objects were used as stimuli, instead of arbitrary patterns. For example, with a combination of psychophysical and ERP measurements, Rousselet et al. (2002) observed no measurable cost for processing two natural scenes (does each scene contain an animal or not?) compared to a single scene. The same categorization task could be performed easily under dual-task conditions (Li et al., 2002)—those same experimental conditions that prevent observers from discriminating, say a red-green disk from a green-red disk (Braun & Julesz, 1998). This was true even if the natural scene target was a vehicle (Fei-Fei et al., 2005; Li et al., 2002), or if the task involved face gender discrimination (Reddy et al., 2004) or face recognition (Reddy et al., 2006). Thus, when stimuli are taken from natural and familiar object or scene categories, attention matters much less for recognition (Braun, 2003). This could imply that for such categories of objects, a form of automatic binding might exist in the brain.

Whether this form of preattentive binding really supports “true”, high-level object recognition is open to debate. For example, Evans and Treisman have suggested that it might only reflect the parallel detection of feature sets (e.g., yellow + stripes + fur + claws . . . = tiger) without a precise spatial binding (Evans & Treisman, 2005; Treisman & Kanwisher, 1998). As I will argue later, however, it can be demonstrated that this form of binding is spatially limited in extent, i.e., features are not collected over the entire visual field, but within a well-defined region of more than three degrees and less than eight degrees of visual angle (for a stimulus at five degrees of eccentricity). We cannot conclude yet whether even more spatially precise binding might occur beyond this resolution—but this need not be so problematic: In the real world (as opposed to laboratory situations), when all the relevant features are present within a spatially limited region (yellow, stripes, fur, claws . . .), one can bet with almost certainty that the object (tiger) is present. If one is reluctant to equate these object representations that are formed in the absence of attention with “true, high-level objects”, then it might be a good compromise to think of them as “proto-objects”, as proposed by Rensink (2002).\(^3\)

---

\(^2\) The term “natural” is used here to refer to ecologically valid stimuli for which the visual system has been optimized throughout its development, as opposed to more arbitrary stimuli that an observer is only likely to encounter in laboratory situations. According to this definition, city scenes, photographs of vehicles, and manmade objects, are all “natural” stimuli.

\(^3\) According to Rensink, proto-objects are “rapidly generated, volatile structures that contain local estimates of scene structure”. They can be “quite sophisticated”, but have “limited spatial coherence”.

The efficient, effortless processing of natural scenes and objects observed under dual-task conditions is a very robust phenomenon: Many recognition and categorization tasks, at different levels of discrimination, can be performed with seemingly little attention. For example, Reddy et al. (2004) observed that face gender discrimination, using faces where obvious external features (e.g., hair, make-up) had been removed, could still be done in dual task. In order to test an even more stringent level of discrimination, they asked observers to recognize a particular male (or female) individual, with distractors being other male (or female) individuals: Still no significant performance decrement occurred when attention was distracted (Reddy et al., 2006)!

**THE LIMITS OF EFFORTLESS PROCESSING**

On the other hand, a limit to this effortless recognition ability can be observed easily, by placing multiple simultaneous stimuli on the screen. We have already described the experiment by Rousselet et al. (2002) in which two natural scenes could be processed in parallel (for an animal vs. nonanimal discrimination). But in another study where four scenes had to be processed simultaneously, performance started to decrease quite strongly (Rousselet, Thorpe, & Fabre-Thorpe, 2004). In fact, when a whole visual search array of natural scenes (up to 16 distinct elements) was presented to an observer, performance (both in terms of reaction time and proportion correct) greatly decreased with increasing set size, i.e., the number of elements on the screen (VanRullen et al., 2004). This indicates an inefficient “serial” search process, and would generally be interpreted as reflecting the involvement of attention in scene recognition—although other interpretations of serial search have been advanced (Carrasco & Yeshurun, 1998; Eckstein, 1998; Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer, 1995; Palmer, Verghese, & Pavel, 2000).

These results are very much in line with classic visual search experiments, which to a large extent have failed to show efficient “parallel” search for line drawings of objects—even natural and familiar ones like animals, vehicles, faces (Wolfe, 1998; Wolfe & Bennett, 1997). Some authors have suggested that faces constitute a special class of stimuli, so important ecologically that they could escape the need for serial search (Hansen & Hansen, 1988; Hershler & Hochstein, 2005). But it has later been shown that these “pop-out” effects were mere reflections of spurious low-level differences between the stimuli used⁴ (Brown, Huey, & Findlay, 1997; Nothdurft, 1993; Purcell, Stewart, & Skov, 1996; VanRullen, 2006). Whatever the type of object involved (even extremely

---

⁴ This inability of faces to “pop out” does not mean, however, that face stimuli might not be special in other regards.
natural and familiar ones like faces), object recognition in visual search thus appears to be a serial process. Therefore, the visual search paradigm yields the opposite conclusion as before: Attention might be necessary for natural scene and object recognition after all—and so presumably for feature binding.

COMPETITION TO THE RESCUE

How can the same discrimination task (e.g., animal vs. nonanimal scene, male vs. female face, ...) be “preattentive” in dual-task, but not “parallel” in visual search? Is attention necessary for binding features into natural and familiar object representations, or not? To answer this, let us consider what should happen within the classic computational approach, based on feedforward and “hardwired” feature combinations (see Figure 1, left side). The common architecture in these models mirrors that of the visual system (Felleman & van Essen, 1991) and involves a succession of processing layers in which neurons combine their inputs within a small region (the “receptive field”) to represent new features of increased complexity: Small luminance patches are combined to represent oriented bars; bars are combined to represent contours; and so on until “recognition” units at the last level respond selectively to particular object categories (Fukushima & Miyake, 1982; Riesenhuber & Poggio, 1999b; Selfridge, 1959; Serre et al., 2007). Increases of representation complexity are generally accompanied by an increase of receptive field size (as observed in the visual system) (Boussaoud, Desimone, & Ungerleider, 1991; Gattass, Gross, & Sandell, 1981; Gattass, Sousa, & Gross, 1988) in order to provide the network with some spatial invariance.

How would such a model perform in the various experimental paradigms described earlier? Because these feedforward models generally don’t include an attention mechanism, dual-task conditions would not change anything to the network behaviour: Recognition would occur as usual, whether or not attention is “distracted”. But when a whole visual search array is presented, features corresponding to several distinct objects can fall within each receptive field, preventing the “recognition” neurons from working properly: Parallel search fails, and attentional requirements emerge (Mozer & Sitton, 1998). This is also a form of the binding problem, i.e., a difficulty in assigning the correct features to the correct objects, as proposed in classic theories; but the main difference is that here binding is only a local problem: It occurs—or fails to occur—within the receptive fields of the recognition units, because of interference and competition between the inputs. Similar recognition failures are commonly observed in single-unit physiology, when both an effective and an ineffective stimulus are simultaneously presented within a V2, V4, or IT
neuron’s receptive field: The resulting neuronal response is unselective and intermediate (Luck, Chelazzi, Hillyard, & Desimone, 1997; Moran & Desimone, 1985; Reynolds, Chelazzi, & Desimone, 1999). Drawing attention to one of the stimuli is required to restore the neuronal response, and alleviate the conflict between the inconsistent inputs. Comparable observations have been made in humans using fMRI (Kastner, de Weerd, Desimone, & Ungerleider, 1998) and ERPs, leading Luck and his colleagues to propose an “Ambiguity Resolution” theory of attention (Luck, Girelli, McDermott, & Ford, 1997). In this view, attention would not be necessary for the binding of features per se (and indeed, would not be needed when a single object is to be recognized), but it would only be critical for resolving the competition that ensues (e.g., in visual search situations) from having multiple items in one receptive field (Desimone & Duncan, 1995; Luck & Ford, 1998; Reynolds & Desimone, 1999).

If receptive field competition is truly the main limitation of parallel processing for natural objects and scenes, then the ability to process simultaneous objects should depend on the distance between them: When they fall in a single receptive field, binding should fail; when they fall in separate receptive fields, (parallel) binding should succeed. This is what we proceeded to verify experimentally.

Dual task with two simultaneous stimuli

First, we adapted the dual-task paradigm in order to evaluate the effects of a simultaneously presented, irrelevant item on object recognition without attention (VanRullen, Reddy, & Fei-Fei, 2005). The attention-demanding task that we used for manipulating the availability of attention was a letter search task (singleton present/absent among five randomly rotated Ls or Ts at the centre of the screen), as used in most of the dual-task literature (Braun, 1998; Braun & Julesz, 1998; Lee et al., 1999; Li et al., 2002; Reddy et al., 2004; Reddy et al., 2006). The task of interest was an animal vs. nonanimal natural scene categorization task, or in other sessions an upright vs. inverted face discrimination task. The stimulus of interest was presented at a random peripheral location (5.5 degrees eccentricity), and in some of the trials a second, irrelevant stimulus of the same type (i.e., a natural scene containing an animal, or an upright face, depending on the session) was shown simultaneously (2.5 degrees eccentricity). The subjects were instructed to ignore this irrelevant stimulus, and in order to make this easier we maintained its identity constant throughout the entire experiment (i.e., for several hundred trials). When this distracting stimulus was present, it was either (with 50% probability) shown close to the target stimulus (3 degrees
away), or far from it (8 degrees away, with the same 2.5 degrees eccentricity but on the other side of fixation). We obtained identical results for the natural scene categorization and the face discrimination task. When the target stimulus was presented alone, it was recognized easily (at around 90% of the baseline single-task performance), even when attention was distracted by the central letter task; this merely confirms previous reports (Fei-Fei et al., 2005; Li et al., 2002; Reddy et al., 2004; Reddy et al., 2006; VanRullen et al., 2004). The presence of the distracting stimulus had a strong impairing effect on dual-task performance (down to \( \frac{65}{2} \), with chance level at 50%), but only when it was shown in close proximity to the target—the remote distracting stimulus had virtually no effect on dual-task performance! This indicates that a binding problem (defined here as interference between two simultaneous items, compared to a situation with a single item) had occurred between the target and the distracting stimulus, but only within a local neighbourhood of the target; a remote distracting stimulus did not trigger this binding problem, and did not impair recognition. Note, finally, that this conclusion is only valid for our natural object or scene recognition tasks: For other, more arbitrary stimuli (e.g., conjunctions of two colours or two orientations), recognition simply cannot occur when attention is distracted, even for a single isolated stimulus (Fei-Fei et al., 2005; Li et al., 2002; VanRullen et al., 2004); it would make little sense in this case to test recognition with two simultaneous stimuli.

**Comparison task**

In order to confirm and extend these results, we designed another paradigm that would allow us to test attentional requirements and interstimulus distance effects for both natural/familiar and arbitrary/artificial object categories in a comparable way (VanRullen et al., 2005). This paradigm involved the comparison of two simultaneously presented items from a given category: Two natural scenes (animal vs. non-animal), two faces (upright vs. inverted), two colour/colour conjunctions (red-green vs. green-red), or two orientation/orientation conjunctions (rotated L vs. T). The subjects responded by indicating whether the items were of the “same” or of a “different” type (e.g., two Ls, or one L and one T, respectively). Critically, the stimuli were masked and the presentation times (stimulus onset asynchrony, SOA) were determined based on pilot data so that identification performance of a single, isolated stimulus was comparable for all stimulus categories (\( \sim 85\% \) correct). We found that observers could not compare the two arbitrary conjunction stimuli (i.e., rotated Ls and Ts, or red-green vs. green-red disks), for which the
comparison performance was near chance (~55%), irrespective of the distance between the two simultaneous stimuli. This confirms the idea that attention is an unconditional requirement for binding and recognition of this type of object. On the other hand, the comparison task was easy (i.e., close to 70%, as expected if both stimuli were identified with ~85% probability) for the natural object and scene discriminations (animal vs. non-animal; upright vs. inverted face) when the distance between the two stimuli to be compared was 8 degrees of visual angle. This confirms that successful binding can sometimes occur “in parallel” for natural stimuli. But when the interstimulus distance was decreased to 3 degrees, a strong decrement was observed in the comparison performance: This reflected a local binding problem, which we attributed to receptive field competition within the relevant neuronal populations.

Visual search with varying interstimulus distance

To finish, we applied our conclusions to the conundrum raised by visual search for natural objects and scenes. If recognition of these stimuli can (in theory) be done preattentively, but is prevented during visual search by unwanted receptive field competition, then visual search performance should be increased by separating elements within the visual search array—as demonstrated in one previous study (Cohen & Ivry, 1991). Furthermore, this should only be true for natural/familiar object categories (because for them the serial nature of the search is merely due to competition), and not for arbitrary/artificial objects (because for them the serial nature of the search reflects a true attentional need). We tested these predictions by evaluating visual search performance for an upright versus inverted face discrimination and a red-green versus green-red disk discrimination tasks, with varying interstimulus distances (Reddy & VanRullen, 2007). The two tasks are similar, in the sense that distractors are simply obtained by rotating the targets by 180 degrees. To avoid eccentricity confounds, all stimuli were presented on a virtual ring at 1.5 degrees eccentricity. At the smallest interstimulus distance tested (0.75 degrees), both tasks showed a clear and similar serial search behaviour (i.e., performance decreased and RT increased significantly with increasing set size). But when interstimulus distance was increased progressively (up to 3 degrees), search performance increased only for the face discrimination task, while no effect was observed for the two-colour disk discrimination—exactly as we predicted. We also verified, under the same experimental conditions, that classic pop-out search tasks (red vs. green patches, + L vs. T symbols) were not affected by interstimulus distance.
Distinguishing preattentive competition from local attentional interference

At first sight, the results described earlier using the dual task, the comparison task, or the visual search paradigm might seem to merely corroborate the numerous previous reports of a spatial pattern of local suppression around the (facilitatory) attentional spotlight (Bahcall & Kowler, 1999; Cutzu & Tsotsos, 2003; Hopf et al., 2006; McCarley, Mounts, & Kramer, 2004, 2007; Mounts, 2000a, 2000b; Mounts & Gavett, 2004; Muller & Kleinschmidt, 2004). This suppressive pattern has been dubbed “localized attentional interference” (Mounts, 2000a, 2000b). In fact, the phenomenon that we demonstrated is quite different: First, in our case the local interference occurs preattentively (as revealed in the dual-task experiment earlier); second, it occurs only for certain familiar and natural object categories, but is absent in other cases (as shown earlier in the comparison task and the visual search experiment). These two properties would not be expected if the interference reflected the suppressive surround of the attentional spotlight; rather, this interference is more likely to reflect competition within the receptive fields of neuronal populations that are preattentively activated. Whether attention, when it finally comes into play, is also accompanied by local suppression, is an interesting question, but one that is orthogonal to our present concern.

TWO TYPES OF BINDING IN THE BRAIN

Putting all these observations together, we suggest that two types of binding might coexist in the brain: Hardwired binding for familiar, natural object categories, and on-demand binding for more arbitrary, artificial, or meaningless conjunctions of features (Figure 1). The former can occur without involving attention, while the latter is directly mediated by attention.

Hardwired binding corresponds to the operation of a system based on a feedforward hierarchy of neurons, which construct their increasingly complex selectivities by pooling the features represented at previous levels. This idea is directly inspired by the organization of the visual system revealed by primate electrophysiology (Barlow, 1972; Felleman & van Essen, 1991; Hubel & Wiesel, 1968). The system is flexible in the sense that it can learn to represent new objects and features over the course of repeated presentations; but for a given stimulus presentation the behaviour of the system is fixed, and fully determined by the existing connectivity between neurons (i.e., it is “hardwired”). Such a system suffers from two main limitations: It can only represent familiar, frequently encountered stimulus categories (and this is probably best if one is to avoid combinatorial explosion problems); and it will fail when the features corresponding to multiple different objects are too close
in space (as in a visual search experiment), since they will clutter the relevant receptive fields. However, notwithstanding these two caveats, such systems can be very powerful since they have the potential to represent multiple objects simultaneously, as long as they are well separated (as shown for example in the “comparison” task), and can function without involving costly attentional resources (as revealed by the “dual-task” paradigm). Indeed, most

Figure 1. Two types of binding can coexist in the brain. One is automatic and preattentive, based on a hierarchy of hardwired conjunctions of features of increasing complexity, as observed in the primate visual system, and described in classic feedforward models of object recognition. It is limited to the processing of familiar, natural object and scene categories. In this example, adapted with permission from Serre et al. (2005), orientation is the only low-level feature depicted, but other features such as colour, motion, binocular disparity, and so on, can also participate in the early, low-level representation. The other type of binding allows the visual system to deal with previously unknown, unfamiliar or arbitrary conjunctions of features by a costly, on-demand binding process mediated by visual attention. As in the classic Feature Integration Theory of attention, this process results in the formation of a temporary high-level representation or “object file”. To view this figure in colour, please see the online issue of the Journal.
state-of-the-art object recognition models are based on this simple idea (Fukushima & Miyake, 1982; Riesenhuber & Poggio, 1999b; Serre et al., 2005; Serre et al., 2007; VanRullen & Thorpe, 2002).

On-demand binding is the form most commonly advocated by psychologists, based on the famous Feature Integration Theory (Treisman & Gelade, 1980). It is the process by which attention combines the various features of a stimulus to construct, one at a time, temporary high-level representations (or “object files” in Treisman’s terminology) that will serve as the basis for recognition, categorization, or behavioural responses (and, potentially, conscious perception). The main amendment that we would like to propose is that this costly process is not required for familiar objects.

A corollary of this distinction is that visual attention serves at least two distinct purposes in the brain (VanRullen et al., 2004). Within the context of a hardwired binding system, attention can be called upon to resolve competition within neuronal receptive fields, a role compatible with the “biased competition” (Desimone & Duncan, 1995; Reynolds & Desimone, 1999) and the “ambiguity resolution” theories of attention. Within an on-demand system, attention serves as the medium that binds features into higher level objects. Whether a unique attentional system accounts for both these duties is an open question, but parsimony would suggest that two distinct structures constitute an unnecessary expense. In any case, while the postulated “hardwired” recognition system easily maps onto the known hierarchy of cortical areas and feedforward neuronal selectivities, the biological source(s) and mode(s) of operation of attention is (are) still largely unknown. There is general agreement, however, from both theoretical and experimental perspectives, that attention heavily involves feedback and reentrant processes to resolve competition and/or allow binding to take place (Bullier, 2001; Di Lollo, Enns, & Rensink, 2000; Hamker, 2003; Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000; Tsotsos et al., 1995). Such a distinction between feedforward hardwired binding and reentrant on-demand binding would then be reminiscent of the separation between “base grouping” and “incremental grouping” postulated by Roelfsema (2006; see also Ullman, 1984).

**CONCLUSION**

It is important to acknowledge that none of the individual ideas expressed here is fully novel or revolutionary. In particular, the present proposal draws on recognized concepts such as Biased Competition and the Ambiguity Resolution Theory, the Feature Integration Theory, and hierarchical feedforward models of object recognition. No alternative is favoured over another, but I have focused instead on revealing the experimental conditions that tend to promote one or the other mechanisms, and have proposed what
seems to be an optimal compromise between two classic, often opposed views of binding. It is thus possible to reconcile computational approaches supporting hardwired binding of natural, frequently encountered object categories, and psychological theories advocating attention-based on-demand binding of arbitrary or meaningless feature conjunctions.

REFERENCES


