Neuron, Vol. 24, 11–17, September, 1999, Copyright ©1999 by Cell Press

The Psychophysical Evidence for a Binding Problem in Human Vision

Review

Jeremy M. Wolfe*‡ and Kyle R. Cave†
*Center for Ophthalmic Research
Harvard Medical School
Boston, Massachusetts 02115
†Department of Psychology
University of Southampton
Highfield
Southampton SO17 1BJ
United Kingdom

What Is "Binding" and Why Might It Be a Problem?

Imagine that you are looking at two women. One has an oval face with striking green eyes framed by long blond hair. The other has a round face with piercing blue eyes framed by wavy red hair. Long before we reach the realms of social psychology, several potential problems present themselves to the visual system. Did those blue eyes go with the blond hair? Was that blond hair wavy? If one woman is Lynn and the other is Anne, which is which? Coherent perception of even a single object requires that the properties of that object be coordinated or *bound* together. As discussed elsewhere in this issue of *Neuron*, information about these properties appears to be distributed across many different brain areas. This separation of different types of information about a single object raises the possibility of a "binding problem." This paper will review some of the psychophysical evidence indicating that this is a real problem that is faced and, under most circumstances, solved by the visual system. We will also discuss contrary evidence that suggests that the visual system has no such binding problem. Finally, we will provide a theoretical framework within which to understand these apparently contradictory data. (Other issues like texture grouping and contour completion might be considered to be examples of binding. In this paper, however, we are restricting ourselves to the binding of features to objects.)

Evidence that There Is a Binding Problem

Illusory Conjunctions

Some of the most striking evidence for a binding problem in vision comes from a class of apparent misperceptions labeled "illusory conjunctions." When subjects must report on the identity of items in briefly presented arrays of colored shapes, they often report seeing a stimulus made up of the color from one array element and the shape from a different array element. Apparently, perceptual features can become unbound from their original objects and can be recombined to form a new object representation.

In Treisman and Schmidt's (1982) classic version of the illusory conjunction paradigm, subjects viewed a line of colored shapes or letters, flanked by two black digits (see Figure 1). Treisman and Schmidt told the subjects that their primary task was to report the two digits, so that their attention would be diverted from the colored shapes. Subjects were very accurate in reporting the digits, but their reports of the stimuli between the digits included a large number of illusory conjunctions. Illusory conjunctions occurred with both letters and abstract shapes and included all the features tested (color, shape, size, and solidity). Treisman and Schmidt concluded that when attention is not available to combine features correctly, they can be put together to form combinations not actually present in the stimulus.

Illusory conjunctions have arisen in a variety of different visual tasks. For instance, Prinzmetal (1981) presented subjects with arrays of circles, with two of the circles each containing a single horizontal or vertical line. Although the lines were at different locations, subjects sometimes perceived them as forming a plus sign. Interestingly, the two lines were more likely to combine into an illusory plus sign if they were both part of the same perceptual group of circles. Prinzmetal and Millis-Wright (1984) demonstrated a different grouping effect in a task in which subjects searched for one of two target letters in an array and then reported its color. Subjects' reports included more illusory conjunctions when the letters formed a word or a pronounceable nonword than when the letter string was unpronounceable. Treisman and Paterson (1984) found that some subjects perceived an arrow when presented with arrays containing the appropriate shape components (a line and an angle). Some of their subjects could also combine an angle and a line into an illusory triangle, but only if circles were also present in the array, presumably to supply a closure feature (for additional demonstrations of illusory conjunctions, see Prinzmetal et al., 1986; Briand and Klein, 1987; Cohen and Ivry, 1989).

While illusory conjunctions have been demonstrated with a number of different methods by a number of different experimenters, there is disagreement over what these errors tell us about visual processes and representations. In the original formulation of Treisman's Feature Integration Theory (Treisman and Gelade, 1980), illusory conjunctions were taken as evidence that binding features into the representation of an object required attention. Preattentively, features were somehow "free floating" and, consequently, capable of arbitrary rearrangement if attention was diverted.

Various aspects of this position have been challenged. For example, Tsal (1989) argued that features could be correctly conjoined without attention and the presence of attention might not always assure correct feature combinations. He and others have argued that illusory conjunctions might reflect a coarse coding of some aspects of feature information (Cohen and Ivry, 1989; Prinzmetal and Keysar, 1989; Ashby et al., 1996). A second important point raised by Tsal and echoed elsewhere is that illusory conjunctions might be failures of memory as much as failures of vision. In a standard illusory conjunction display, the stimuli are presented briefly and the subject is asked to describe what was seen (but see Prinzmetal et al., 1995).

‡To whom correspondence should be addressed (e-mail: wolfe@search.bwh.harvard.edu).

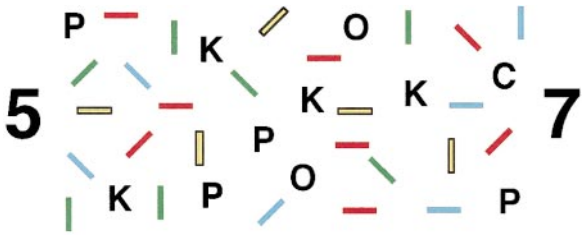

Figure 1. Illusory Conjunction Demonstration

(1) There are two large numbers on the left and right of this figure. Determine if they are both odd and then read ahead for more instructions.

(2) Without looking back at the figure, ask yourself if it contains the letters "R," "P," or "Q"? Did you see a vertical yellow bar? Did you see a horizontal green bar? If you thought you saw an "R," a "Q," or a horizontal green bar, you have made an "illusory conjunction" error.

In a somewhat similar vein, Butler, Mewhort, and Browse (1991) claimed that illusory conjunction errors demonstrate more about the encoding strategy used in a particular task than about the basic properties of visual representations. They showed that the same stimuli could produce different patterns of errors depending on subjects' expectations. When subjects knew that they would see only uppercase letters in each trial, they would sometimes combine a bar from a letter Q with a letter P and report a letter R. However, when subjects did not know whether to expect upper- or lowercase letters, then their errors with the same uppercase stimuli consisted mainly of mislocating entire letters rather than combining features from different letters. Butler, Mewhort, and Browse concluded that subjects encode the stimuli as features in the first task and as letters in the second task.

Illusory conjunctions are not limited to basic preattentive features: just as features can travel from object to object, letters can travel from word to word (Mozer, 1983; McClelland and Mozer, 1986). When Mozer's subjects viewed the two words "LINE" and "LACE," they sometimes reported seeing "LICE" or "LANE." As in the earlier demonstrations of illusory conjunctions, these letter migrations occurred more often when attention was diverted to other stimuli. They were also more likely to occur when the two words shared some letters, and they occurred just as often whether the letters in the two words matched in case or not, indicating that the confusion occurred within abstract word representations rather than between representations of individual letters or features (see also Treisman and Souther, 1986; Fang and Wu, 1989). Even more abstract errors appeared in experiments by Virzi and Egeth (1984), in which subjects confused the color named by a word and the color of the ink with which it was written (see also Intraub, 1985; Goolkasian, 1988).

To summarize, illusory conjunctions are clear evidence for *some sort* of problem with the correct binding of features to objects. The original conception of "free floating" preattentive features bound together by the glue of attention has been replaced by a view that illusory conjunction phenomena can occur when linkages break down at any of a number of levels of processing. These linkages may be built and maintained by attention, at least at some levels, thus preventing illusory conjunctions.

Dissociating Detection and Localization of Features

If binding is a problem, then it should be possible to find evidence of unbound features. What would such evidence look like? If some process like attention is needed in order to associate features with the correct locations, then it should be possible to dissociate the identification and localization of features. The strong form of this hypothesis would hold that there exists a preattentive stage of processing at which features are identified but represented completely independently of location. Original Feature Integration Theory argued for a position of this sort. Treisman and Gormican (1988), for example, argued that detection of features and localization of features were separate operations, though Treisman's more recent views are less absolute (Treisman, 1996). Still, one of the main claims of Feature Integration Theory (Treisman and Gelade, 1980) and related models like Guided Search (Wolfe, 1994a) is that the features belonging to visual objects cannot be accurately bound together into object representations in the early, preattentive stage of visual processing.

Treisman and Gelade (1980) originally predicted that subjects performing a feature search might be able to report the presence of a target feature in a display even if they were unable to localize the target. The feature would be detected preattentively without the localizing effects of spatial attention. In conjunction search, however, this dissociation would not be possible because conjunctions could not be detected without spatial attention. Thus, if a subject could detect a conjunction target, then they should also be able to localize it. Treisman and Gelade tested their prediction in two experiments in which each search array contained one of two possible targets. In the *feature search* condition, the target had a color or shape not shared by any of the distractors. Here, accuracy in reporting the target feature was above chance, even on those trials with large errors in the report of the target location. However, in the *conjunction search* task, subjects did not report the conjunction features accurately unless they also correctly reported the location of those features.

Nissen (1985) provided additional evidence that color and shape features from an object were combined via location. She presented subjects with an array of four objects, each with a unique shape and unique color. In one condition, one of the four possible locations was cued before each trial, and the subject reported the color and shape at that location. Nissen predicted that accuracy of the color and shape reports should be independent of one another, because subjects did not need to determine one feature in order to determine the other. The data were consistent with that prediction. In a second condition, a color was cued at the beginning of each trial. Subjects were asked to report the location and the shape of the object with the cued color. In this condition, Nissen predicted that the accuracy of the shape reports would depend in part on the accuracy of the location reports, because subjects would have to determine the target location before they could determine its shape. The data showed that the shape and accuracy reports were not independent, and that shape

accuracy was very low when the location was reported incorrectly. From the first condition, Nissen estimated the proportion of targets for which the shape would be correctly reported once the location was known. Using this estimate and the proportion of correctly reported locations in the second condition, she was able to work out predictions for the second condition that corresponded fairly closely to the actual results.

Taken together, the Treisman and Gelade (1980) and Nissen (1985) results suggested that features could be represented without being bound to their locations, so that subjects could report feature identities without locations. Subsequent work, however, has raised questions about these free-floating features. Johnston and Pashler (1990) questioned whether Treisman and Gelade were correct in concluding that features could be identified without being located. They suggested that even if a feature's location had been correctly determined, subjects might not always be able to report its position accurately using Treisman and Gelade's system for reporting location. They also conjectured that accuracy in reporting features may have been deceptively high in Treisman and Gelade's experiment, because when subjects were unable to detect either of the possible target features, they might guess and report the target that was more difficult to detect (a "negative information" strategy). Johnston and Pashler performed their own version of the experiment in which the stimulus elements were arranged so that each location occupied a unique corner or side, making it easier to remember and report each location. They also tried to equalize the discriminability of each of the two target features (although they concluded from their results that they were only partially successful). They found only weak evidence for identification without localization and concluded that the phenomenon was rare, at best.

Just as Johnston and Pashler (1990) raised doubts about Treisman and Gelade's evidence for unbound features, a later study by Monheit and Johnston (1994) raised doubts about Nissen's claims for the independent reporting of color and shape. With a careful analysis of Nissen's task, Monheit and Johnston demonstrated that any effects of nonindependence between color and shape would be very small if the subjects used a reasonable guessing strategy when they failed to identify the features present. They then conducted their own versions of Nissen's experiments with some changes to increase their ability to detect nonindependence effects. They increased the number of trials per subject, and they selected their stimulus elements from a set of six colors and six shapes (rather than the four used by Nissen) to limit the effects of guessing. They found the nonindependence effects they expected: in many trials, subjects either reported both the color and shape correctly or got them both wrong.

The experiments discussed here are only a small portion of the literature on the independence or lack of independence between identification and localization tasks. The Treisman and Nissen studies argue for a precedence for identification over localization. Other studies have argued for the opposite (Sagi and Julesz, 1985; but see also Folk and Egeth, 1989) or for an equality between the operations (Green, 1992). Saarinen has argued that it is futile to search for a clear answer in this line of research (Saarinen, 1996a, 1996b). The area suffers from a pair of seemingly insurmountable problems. First, accuracy and reaction time data, the measures of choice, can be readily altered by task manipulations. Second, data about the state of vision *prior* to the deployment of attention are derived from measures taken *after* the stimulus is gone, just as in the illusory conjunction experiments. This makes it hard to know if a failure to report the locus or identity of a stimulus is the result of a failure to process or a failure to remember. We cannot be sure if subjects misperceive a combination of two features or misremember that combination.

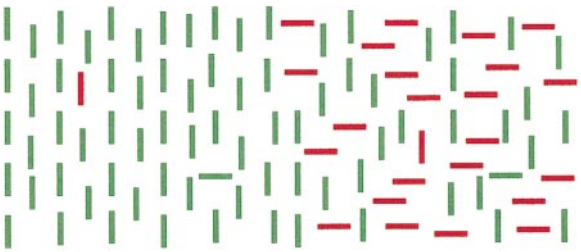

Figure 2. Feature and Conjunction Search

It is very easy to find the red and vertical items on the left of this figure. On the right, the item defined by the conjunction of red and vertical does not "pop out" in the same way.

To summarize this section, there is some evidence for a dissociation between identification and localization of basic feature information. For present purposes, this evidence supports the notion that there is a binding problem in early vision. However, interpretation of these data has proven to be ambiguous and the experiments, taken as a whole, make for a somewhat unsatisfying meal. Clearer evidence that the visual system faces problems in binding features into object representations comes from the visual search literature.

Search for Conjunctions of Basic Features

One of the pillars of support for the existence of a binding problem in human vision was the apparent inefficiency of searches for targets defined by conjunctions of basic features. A search for a target among distractors is very easy if the target is defined by a single salient feature. Thus, as shown in Figure 2, it is easy to find the red item among green distractors or the horizontal item among vertical distractors. However, Treisman and Gelade (1980) reported that the same features failed to produce efficient search when those features conjunctively defined the target (see the right side of Figure 2). They found that a search for a red vertical target among red horizontal and green vertical distractors produced an inefficient search, consistent with a serial, self-terminating search through the items. In the original Feature Integration Theory, this apparent seriality was taken as evidence that features were unbound prior to the arrival of attention.

Subsequent research complicated the picture from the vantage point of the binding problem. Houck and Hoffman (1986) demonstrated that unattended conjunctions of color and orientation could produce a McCollough effect, a visual aftereffect dependent on the contingent relationship between color and orientation. Even more troubling was the data from numerous labs showing efficient search for conjunctions (e.g., Nakayama

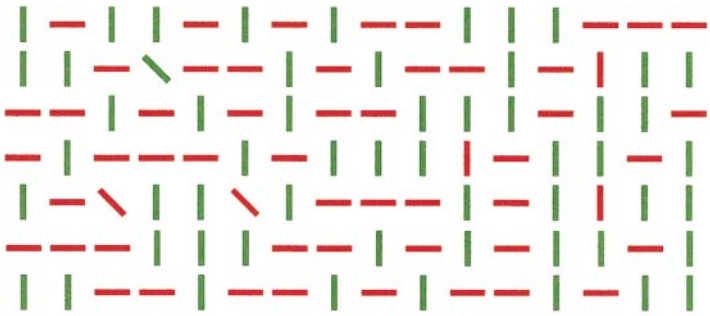

Figure 3. Texture Segmentation by Feature and Conjunction

In the left half of the figure, it is quite easy to see the triangle formed by the three oblique elements. On the right, the three red vertical elements never form a triangle of comparable clarity.

and Silverman, 1986b; Quinlan and Humphreys, 1987; Wolfe et al., 1989; Zohary and Hochstein, 1989; Treisman and Sato, 1990; Cohen and Ivry, 1991; McLeod et al., 1991). At first, it appeared that these results might be specific exceptions to the general rule of inefficient conjunction search (Nakayama and Silverman, 1986a; McLeod et al., 1988). However, subsequent work indicates that *any* search for conjunctions of basic features is efficient if the features are salient enough (see discussions in Wolfe, 1994a, 1998). Indeed, there are several published reports of conjunction searches that yield search efficiencies that are indistinguishable from those produced by basic features (e.g., Wolfe, 1992; von der Heydt and Dursteler, 1993; Theeuwes and Kooi, 1994).

Do these findings argue against the existence of a binding problem? Do they show that features are conjoined prior to the arrival of attention? Perhaps not. Efficient search for conjunctions can occur even if features cannot be bound together without attention. Consider the search for a red vertical item among green vertical and red horizontal items. Suppose that a parallel color processor biases the deployment toward red items and an orientation processor biases the deployment of attention toward vertical items. Even though color and orientation are being handled entirely separately, the combination of these two sources of attentional *guidance* will tend to deploy attention to loci containing both red and vertical. This concept of guidance is, not surprisingly, at the heart of the eponymous Guided Search model (Wolfe et al., 1989; Cave and Wolfe, 1990; Wolfe, 1994a; Wolfe and Gancarz, 1996). It is also a part of later versions of Feature Integration Theory (e.g., Treisman and Sato, 1990) and the more recent FeatureGate model (Cave, 1999) and is anticipated by the work of Hoffman (1979; see also Tsotsos et al., 1995).

These developments in the understanding of conjunction search render ambiguous the role of attention in conjoining features. Is item-by-item attention needed to bind features into objects or not? More recent conjunction experiments suggest that Treisman's original claim is correct even if the original empirical support for the claim is open to reinterpretation.

The first case, illustrated in Figure 3, is derived from Wolfe et al. (1995). On the left side of the figure, the three items of odd orientation form a virtual triangle that is detected without noticeable effort (Nothdurft, 1992). Wolfe et al. (1995) had subjects describe the orientation of a briefly presented triangle of this sort. The task was easy even when three different orientations formed the vertices of the triangle. However, when the vertices were defined by conjunctions of color and orientation, as they are on the right side of this figure, no impression of a triangle was instantly available. Instead, subjects seemed to need to attend to each red vertical in turn in order to describe the position of the triangle as a whole. In a brief presentation, the task was essentially impossible.

Figure 4 shows a second illustrative case. Wolfe and Bennett (1997) had subjects search for red vertical lines in displays similar to those shown here. On the left of Figure 4 is a typical conjunction search. On the right, the same elements have been combined into "pluses." Search was much less efficient in the latter case. Wolfe and Bennett argued that this case represented conjunction search in the absence of useful guidance. On the left, it is possible to guide attention toward "red" items and toward "vertical" items. On the right, *all* items contain "red" and "vertical." Prior to the arrival of attention, each plus is the same preattentive bundle of "red" and "green" and "vertical" and "horizontal." It is only when attention is deployed to an item that it is possible to correctly bind colors and orientations.

To summarize this section, recent research supports a modified version of Treisman's position on conjunctions. When Feature Integration Theory was first proposed, it was unclear if subjects searched from object to object in the visual field or from location to location. In that context, Treisman could propose that features were initially "free floating." As will be discussed later in this paper, subsequent work has made it clear that search generally proceeds from object to object (e.g., Behrmann and Tipper, 1994; Tipper et al., 1994; Vecera and Farah, 1994; Wolfe, 1994b, 1996; Yantis and Gibson, 1994; Wolfe and Bennett, 1997; Tipper and Weaver,

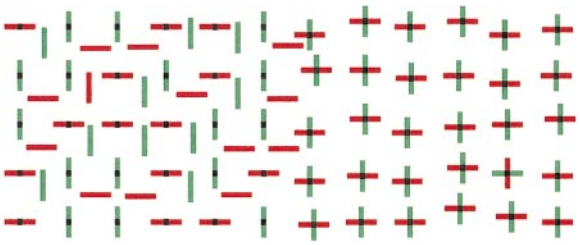

Figure 4. Preattentive Objects Are Just Bundles of Features

Search for red vertical items. On the left side of this figure, the task is a relatively easy "guided" search. On the right, the same red vertical element is very difficult to find because all of the elements contain the features "red," "green," "vertical," and "horizontal."

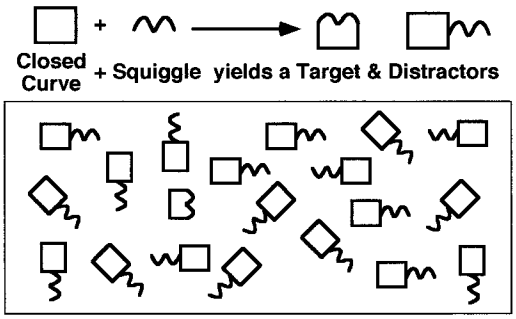

Figure 5. An Example of Inefficient Search for a Shape
From Wolfe and Bennett (1997).

1998) (more specifically, attention appears to be directed to locations defined by objects; reviewed by Cave and Bichot, 1999). In this light, we can describe the preattentive world as populated by unrecognized bundles of features loosely held together by virtue of their shared location. Thus, all of the pluses in Figure 4 are bundles of red, green, vertical, and horizontal. Attention is required to correctly *bind* the features into a red–vertical, green–horizontal plus or into its 90° rotation. Similarly, while objects like faces may be composed of basic features that are processed without attention, it is only with the deployment of attention that these features can be bound together into a representation that can be recognized (Nothdurft, 1993; Suzuki and Cavanagh, 1995; Cave and Bichot, 1999).

Objects as Bundles of Features

The idea that objects are represented only as bundles of basic features prior to the arrival of attention can be used to explain failures to search efficiently for targets defined by the spatial arrangement of their parts. Searches for Ts among Ls and for Ss among mirror-reversed Ss are standards of inefficient search (e.g., Braun and Julesz, 1998; Kwak et al., 1991), though there are some reports of efficient search for targets defined by spatial relations (Wang et al., 1994) or even by their conceptual category (Jonides and Gleitman, 1972; but see Duncan, 1983; Dixon and Shedden, 1987; but then see Krüger, 1984).

The same pattern is seen in objects that are not alphanumeric characters. Target shapes that are quite different from distractor shapes yield inefficient search if they are composed of the same bundle of features. For example, a closed curve and a squiggle, as shown in Figure 5, can be combined to make two very different objects. Nevertheless, because they are both composed of the same preattentive bundle of features, search for one of these among the other is quite inefficient (Wolfe and Bennett, 1997).

Evidence that There Is Not an Absolute Binding Problem

The evidence discussed thus far indicates that the visual system struggles with a binding problem and sometimes loses. Prior to the arrival of attention, the features of an object seem to be rather loosely affiliated with each other. The relationship of color to orientation or squiggle to closed curve seems to be properly appreciated only after attention is deployed to an object. However, as noted above, features are not entirely independent in the absence of attention. Houck and Hoffman (1986), cited earlier, showed that attentional manipulation did not disrupt the McCollough effect, an aftereffect dependent on a firm, contingent relationship of color and orientation (McCollough, 1965). If preattentive features were utterly unbound, it should not be possible to produce an effect that requires the association of, say, red and vertical (Dodwell and Humphrey, 1992).

Like the Houck and Hoffman experiments, there are other studies showing a relationship between features in the absence of attention. We would argue that these studies show that features of an object are *bundled* together preattentively but that explicit knowledge of the relationship of one feature to another requires spatial attention.

Features Linked to Other Features of the Same Object

A number of studies demonstrating object-based attention suggest that all of the features belonging to an object are bundled and selected together. For instance, Treisman, Kahneman, and Burkell (1983) asked subjects to perform two tasks: reading a word aloud and localizing the gap in a rectangle. Performance was better when the rectangle surrounded the word, making them a single object. This advantage could not be attributed to distance, because the distance between the word and the gap was the same whether the rectangle surrounded the word or not.

Further evidence for object-based attention came from another dual-task experiment by Duncan (1984). The stimuli consisted of a diagonal line and a rectangle superimposed. Depending on the condition, subjects reported either two properties of the line, two properties of the rectangle, or one property of each. Reports were more accurate when both properties were from the same object. Because the objects were superimposed, Duncan argued that the same-object advantage could not result from selecting the object's location. Many subsequent studies have produced similar results (for a useful review, see Goldsmith, 1998; see also Tipper et al., 1994; Vecera and Farah, 1994; Lavie and Driver, 1996; Tipper and Weaver, 1998) (although Cave and Kosslyn, 1989, argued that the object is selected by a very specific selection of locations).

Recent evidence about the extent of feature bundling can be found in the experiments by Luck and Vogel (1997). On each trial, their subjects viewed two multielement displays, with a delay of less than a second between them, and monitored one feature dimension for changes between the two displays. With set sizes of up to four, subjects could keep track of the color or shape or size or orientation of four objects without much trouble. Interestingly, performance was just as good when subjects had to look for changes that could occur in any of the four features. Seemingly, subjects were now remembering 16 pieces of information. We know, however, that subjects could not remember the colors of 16 distinct objects, so the results suggest that visual short-term memory can hold approximately four objects and that all of the features of each object are recorded and bundled together.

When Is Binding a Problem?

The psychophysical research reviewed here makes it clear that there is a binding problem in human vision. There are circumstances under which observers behave as if basic features that are tightly linked in the world are, at best, loosely linked in the visual system. At the same time, the orderly and reasonable nature of routine visual perception demonstrates that the visual system solves the binding problem successfully most of the time. By "successfully," in this case, we mean that the problem does not interfere with our usual uses of our visual systems. The visual system may be rife with unbound features but, under normal circumstances, they do not intrude. Why not?

We propose that there are two answers to this question: a preattentive or, perhaps, unattended answer and an attentive answer. In the early stages of visual processing (i.e., primary visual cortex), visual information is represented within spatially organized maps of the visual field. Each feature is represented as occupying a fairly specific location, and thus location serves as a means for linking all of the features belonging to a single object. Although the representations at this level contain all of the information necessary to determine the relationships between the features in an object, those relationships are not explicitly represented at this level. Without explicit representations of the feature combinations, target objects defined by a combination of features cannot be found easily in visual search, although a feature combination may produce some form of priming or adaptation. We can think of the features at this early level as being loosely "bundled" together rather than tightly "bound."

In the absence of visual attention, the spatially organized maps of the visual field would prevent features from becoming truly "free floating." However, without the explicit representation of the relationships among features, or "binding," permitted by the deployment of attention, it may not be possible to recognize these spatially correlated bundles of features. The processes of object recognition require that features be tightly "bound" rather than loosely "bundled," as they were in the earlier levels. The binding, however, is only possible for objects selected by attention, and not for all of the objects present at unselected locations in the visual field.

The simple spatial association used in early vision may not help in later stages of object recognition (i.e., the inferior temporal lobe), because specific information about the location of each feature is no longer available. In order to avoid the combinatorial disaster of representing *all* objects in *all* orientations at *all* locations, cells with complex response properties respond to those properties across large portions of the visual field. If information from multiple objects in the visual field were represented simultaneously at this level, it would be difficult to determine which features belonged to which objects. Selective attention is the apparent solution to this aspect of the binding problem. If visual selection mechanisms allow only selected objects or locations to be represented at this level, then the specific relationships among features can be represented explicitly. In this way, a mechanism that is specialized for face recognition, for example, can receive two eyes, a nose, and a mouth. If attention did not regulate the input into the face recognizer, it might receive six eyes, three noses, and three mouths. In the absence of the tight spatial information of earlier visual stages, it might be quite unable to associate the correct facial features with each other.

Thus, across the great bulk of the visual field, unrecognized objects are held together by the spatial organization of the early visual system. At later stages, a recognized object is held together by the explicit binding of a selected set of features. Working in tandem, these processes of bundling and binding deliver a coherent perceptual world. The data described above suggest that problems in feature binding arise in two circumstances. The first occurs when explicit representations of feature combinations are needed before the objects have been selected. This situation can happen in visual searches for targets defined by feature combinations. We can't easily search for one face in a field of distractor faces with similar features, because the featural configuration for each face cannot be represented until that face is individually selected and its representation is built in the later stages of the visual system. Guidance by basic features limits this problem in most cases. Thus, the search for a face is only noticeably inefficient if there are many faces. Otherwise, basic feature information can guide attention to the few faces in the scene. Similarly, recognition of a specific car in the parking lot requires explicit binding of its features. However, in the search for your yellow Volkswagon (Weisstein, 1973), attention will be guided to items of the appropriate color. The visual search examples that point to failures of binding (e.g., the pluses of Figure 4) must be carefully contrived to require binding without permitting guidance.

The second set of circumstances that produces a binding problem is exemplified by illusory conjunctions. When information is presented briefly or when sustained information is not recoded into memory, the spatial glue that holds bundled features together becomes degraded. Basic features with uncertain positions can combine to produce illusory conjunctions. Recall that the stimuli are gone in most cases of illusory conjunctions, making accurate updating of spatial position impossible.

We make no specific claims here about the neural substrate of binding. This account neither requires nor contradicts a feature binding role for oscillations or some other form of synchronous neural firing, as has been proposed in a number of different contexts (e.g., von der Malsburg, 1981; Crick and Koch, 1990a, 1990b). Whatever the mechanism, we would be surprised if it did not produce an early, parallel bundling of features into objects at specific spatial locations followed by a later selection of one or more of those bundles for more precise binding.

In summary, this paper has described some of the psychophysical evidence for binding problems in human vision. Other papers in this issue deal with the physiological and/or computational solutions to these problems. Psychophysics points to two aspects of those solutions. Early stages of visual processing appear to be able to divide the world into proto-objects that are little more than loosely organized feature bundles at specific locations in space. These initial object parsing operations

are apparently performed in parallel across the visual field and prevent features from floating freely by tying them to spatial locations. Later stages, responsible for object recognition, require tighter, more accurate binding of features and more explicit representations of the relationships among the features. This more demanding stage is capacity limited. Attentional selection is used to restrict this more complete binding to the current object (or objects) of atttention.

Acknowledgments

This work was supported by grants from the National Institutes of Health–National Eye Institute, the National Institutes of Mental Health, the National Science Foundation, the Air Force Office of Scientific Research, and the Human Frontiers Science Program. We thank Bill Phillips for comments on an earlier draft.

References

A comprehensive reference list for all reviews can be found on pages 111–125.