6

Feature binding, attention and object perception

Anne Treisman

Abstract

The seemingly effortless ability to perceive meaningful objects in an integrated scene actually depends on complex visual processes. The 'binding problem' concerns the way in which we select and integrate the separate features of objects in the correct combinations. Experiments suggest that attention plays a central role in solving this problem. Some neurological patients show a dramatic breakdown in the ability to see several objects; their deficits suggest a role for the parietal cortex in the binding process. However, indirect measures of priming and interference suggest that more information may be implicitly available than we can consciously access.

6.1 The binding problem

The binding problem in perception deals with the question of how we achieve the experience of a coherent world of integrated objects, and avoid seeing a world of disembodied or wrongly combined shapes, colours, motions, sizes and distances. In brief, how do we specify what goes with what and where? The problem is not an intuitively obvious one, which is probably a testimony to how well, in general, our brains solve it. We simply are not aware that there is a problem to be solved. Yet findings from neuroscience, computer science and psychology all imply that there is.

There is considerable evidence that the visual system analyses the scene along a number of different dimensions in various specialized modules. Both anatomical and physiological evidence (reviewed, for example, by Cowey (1985) and Zeki (1993)) suggests the existence of several maps of the visual scene laid out in different visual areas of the brain. Recordings from single or multiple neurons in animals have shown different specializations. Ungerleider and Mishkin (1982) distinguished a dorsal pathway, coding motion and space, and a ventral pathway, coding colour, shape and other features in extrastriate areas and eventually objects in area IT. Consistent with this inferred modularity, localized brain damage in human patients leads to selective losses in perceptual abilities. For example, colour vision can be lost in achromatopsia, without any impairment in shape or motion perception (Meadows 1974; Damasio *et al.* 1980); the ability to perceive motion can also be independently lost in akinetopsia, resulting in perception of frozen stills (Zihl *et al.* 1983; Zeki 1991); so can the ability to discriminate orientations or simple shapes (Goodale and Milner 1992). Finally, in humans with intact brains, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have shown focal activity shifting to different brain areas as subjects are asked to respond to different aspects of the same displays—the shapes, colours and directions of motion (see Corbetta *et al.* 1991; Gulyas and Roland 1991; Sereno *et al.* 1995).

These findings, suggesting that specialized areas code different aspects of the visual scene, raise the question of how we get from dispersed brain representations to the unified percepts that we experience. If the world contained only one object at a time, this need not be a problem: there is nothing to demand that a unitary percept must depend on a unitary localized neural code.

However, the binding problem is raised in a more acute form when we realize two facts: first, that we typically do not look at scenes with only one object in them. The world around us is a crowded place, full of objects. Second, receptive fields in many of the specialized visual areas are quite large—up to 30° in temporal areas. Beyond the earliest stages of visual processing, single neurons respond across areas that would certainly hold several objects in crowded displays. If two objects with potentially interchangeable properties are detected by the same units, the potential for miscombining is present. For example, if a unit responding to red is active at the same time as a unit responding to motion, we need some way of distinguishing whether their receptive fields contain a moving red object, or a moving green object together with a stationary red object.

Which mechanisms could resolve this ambiguity? One possibility is that single units directly code conjunctions of features at earlier levels where receptive fields are small enough to isolate single objects. Certainly most cells in both early and late visual areas are selective along more than one dimension. Tanaka (1993) has shown single units in area IT that respond to relatively complex combinations of features. But in these experiments the animals were typically shown one object at a time, so the binding problem did not arise. The cells in IT could be coding the output of the binding process. There must be limits to the use of direct conjunction coding as a solution to the binding problem. We can see an essentially unlimited number of arbitrary conjunctions, immediately, the first time we are shown them. A purple giraffe with wings would look surprising but it would not be invisible. There are certainly too few neurons to code individually the combinatorial explosion of arbitrary conjunctions that we are capable of seeing.

A suggestion that is currently arousing interest in both neuroscience and computer modelling is that binding might depend on synchronized neural activity. Units that fire together would signal the same object. Gray *et al.* (1989) and Singer and Gray (1995) have collected evidence showing the presence of stimulus-dependent synchrony between units in quite widely separated areas of the brain. It is an interesting hypothesis, but I don't think it solves the same binding problem that I raised at the beginning of this paper. Synchrony is a possible way of holding on to the solution, of tagging the units that are responding to the same object once they have been identified, but we still need a way of finding which those are. The Gestalt psychologists identified a number of perceptual cues, such as collinearity, proximity, similarity, which determine perceptual grouping within dimensions such as colour, orientation and common motion. Facilitatory connections between cells responding to the same or related features within dimensions might mediate this grouping by helping to synchronize their firing across different locations (see, for example, Hummel and Biederman 1992), but risk also leading to false bindings when different objects share the same features. Furthermore, they would not bind the different features like orientation, motion, and colour, that happen to belong to the same object. This paper suggests a possible mechanism for binding across dimensions through shared locations, and also for using similarity to bind within dimensions across locations.

6.2   A role for spatial attention?

Psychologists have been interested for many years in a spatially selective mechanism of visual attention. For example, Posner (1980) showed that giving a spatial cue, such as a momentary brightening of one of two frames, would speed responses to a target object that subsequently appeared in that frame, even when the subject's eyes remained fixated centrally. We use the analogy of a 'window' of attention for this unitary, spatially selective mechanism. Other experiments have investigated visual search by asking subjects to find a target object in a display of nontargets (we call them distractors). We measure how long it takes to find the target as a function of how many distractors there are in the display. In some search tasks, the search time increases linearly with

the number of distractors, as though subjects used a serial process of checking objects to find the target. Perhaps the same attention window must be centred on each object in turn. There is evidence that the attention window can be scaled—its size can adjust to fit the objects or groups of objects that are relevant to the task. For instance, in a display containing a global letter made, of smaller local letters, we can attend to the global letter or to any one of the local ones, and it ' takes time to switch between these two states (Navon 1977; Ward 1982). In search, we process homogeneous groups of items in parallel (Treisman 1982).

Some years ago, I suggested that spatially selective attention may play a role in binding (Treisman and Gelade 1980; Treisman 1988). The idea, a very simple one, was that we code one object at a time, selected on the basis of its location at an early level where receptive fields are small. By temporarily excluding stimuli from other locations, we can simply bind whatever properties are currently attended. Figure 6.1 shows the model I proposed to relate the early parallel stages of vision to later attentional stages. It includes a master map of locations, that registers the locations of regions without giving access to the features that define them—for example whether they represent discontinuities in luminance, colour, depth or motion—and a separate set of feature maps. The feature maps contain two kinds of information: a 'flag' signalling whether the feature
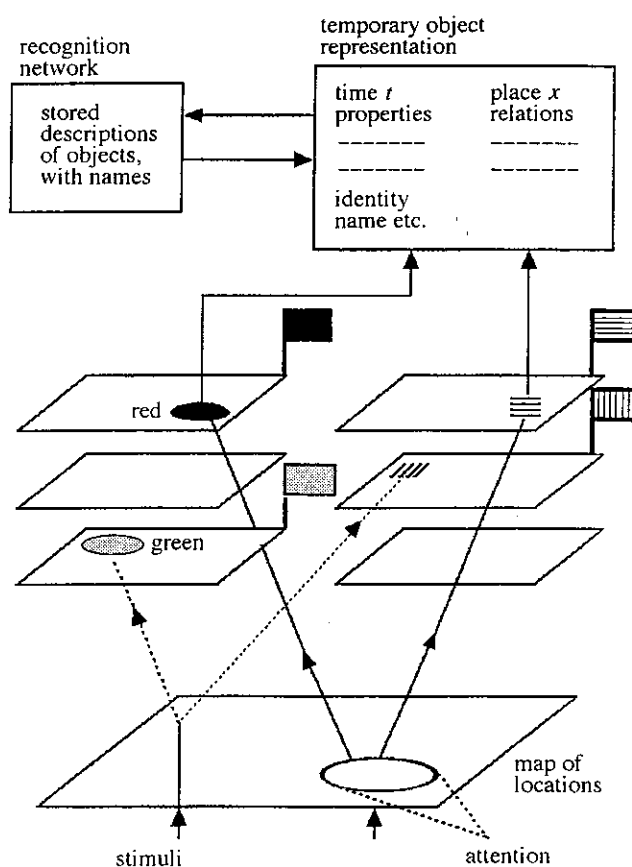

Fig. 6.1   Model suggesting the relation between feature coding, spatial attention and binding in object perception.

is present anywhere in the field, and some implicit information about the current spatial layout of the feature. Not all tasks require binding. If a task can be done simply by checking the flag for the presence of activity within a single feature map, it should not depend on attention. So, for example, the information that there is something red out there can be accessed directly from the feature maps, but the location of the red thing and its other features cannot.

The hypothesis is that locating and binding the features requires retrieval of their connections to the master map of locations. To put 'what' and 'where' together, an attention window moves within the location map and selects from the feature maps whatever features are currently linked to the attended location, temporarily excluding the features of all other objects from the object perception level. The attended features can then be entered, without risk of binding errors, into the currently active object representation where their structural relations can be analysed. At the instant shown in the figure, the information explicitly available would be a detailed specification of the object currently in the attentional window, plus the fact that green and vertical are present elsewhere. There might also be surviving representations of previously attended objects, although, surprisingly, there is some evidence that the bindings are lost as soon as attention is withdrawn (Wolfe 1998). Once a unitary object has been set up, it can be matched to stored models and identified, and actions such as reaching or grasping it can be programmed.

6.3 Evidence from illusory conjunctions

Next, I will outline some behavioural evidence that seems consistent with this hypothesis. Perhaps the most dramatic comes from a patient who seems to have severe problems in binding features (Friedman-Hill *et al.* 1995; Robertson *et al.* 1997). They illustrate what can happen when binding breaks down. We showed R.M. some very simple displays containing just two coloured letters selected from T, X, and O in red, blue, or yellow, and asked him to tell us the first letter he saw (figure 6.2). The exposure durations ranged from 0.5 to 10 s. In some sessions, even with exposures as long as 10 s, he made binding errors, reporting one letter in the colour of the other, in more than 35% of trials. He reported a feature that was not in the display in less than 10% of trials. If he were guessing, these two kinds of errors would be equally likely, as there was always one other colour or shape in the display and one not presented, so we can infer that one-quarter to one-third of his responses were binding errors. Clearly something had gone very wrong with his ability to bind. He had lost the ability that we all rely on to see stable well-integrated objects,
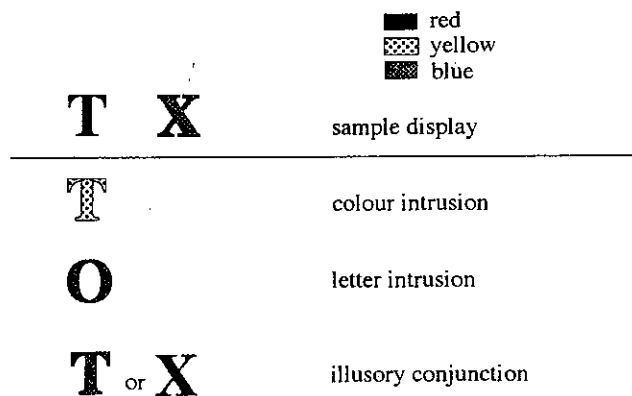

Fig. 6.2 Stimuli used to test feature binding in patient R.M.

and he now lives in a troubling world in which the binding problem is one he must constantly confront. I will return later to discuss other aspects of his perceptual problems.

Are there any conditions in which normal people have similar problems? As the hypothesis was that spatial attention is involved, we tried to prevent people from focusing attention by giving them a brief presentation and requiring them to spread their attention globally over the whole display (Treisman and Schmidt 1982). In one experiment, the displays contained four shapes varying in colour, size, shape, and format (filled or outline) arranged at the corners of a square, flanked on each side by two smaller black digits (figure 6.3a). Subjects were asked to give priority to noting the digits and to report them first. In addition they were to report all the features they could of the shape in one of the four locations, cued by a bar marker which appeared 200 ms after the display. The prediction is that they too should then make binding errors, putting features together in the wrong conjunctions. Subjects did in fact make many conjunction errors recombining the shape,
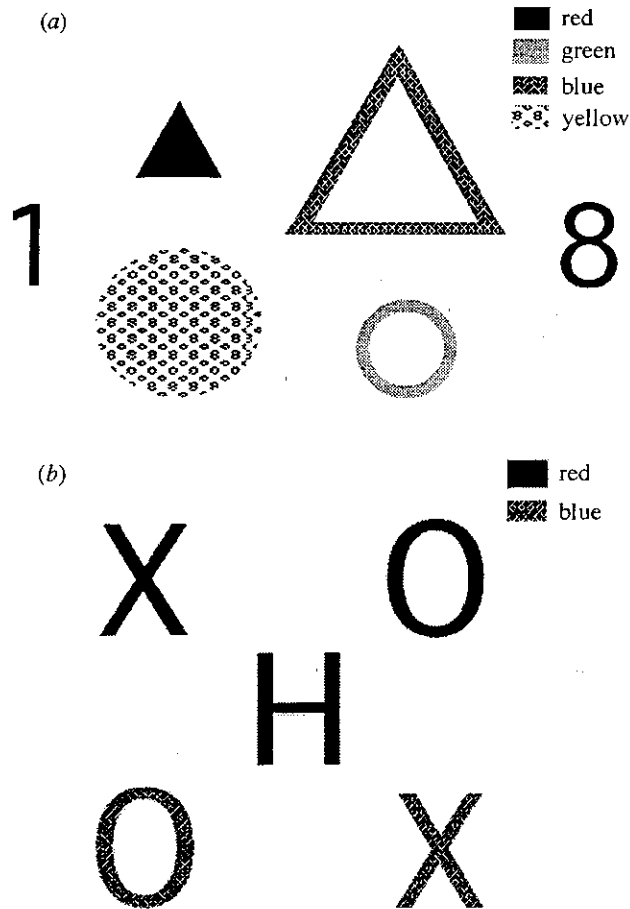

Fig. 6.3   (a) Display illustrating task to test the role of attention in binding in normal subjects. The task is to report the identity of the two digits first and then as many features as are confidently seen for the object in the cued location, indicated by the bar marker. (b) Display illustrating same–different matching task that reveals binding errors without a load on memory.

size, format and colour. These occurred in 18% of trials, compared with only 6% intrusion errors. Again, we would expect equal numbers of each if subjects were guessing or misperceiving the individual features. Instead they seemed to form illusory conjunctions recombining features that in fact characterized separate objects. In another experiment with coloured letters, there were more than 30%. We called these errors 'illusory conjunctions', implying that they are real perceptual illusions. Having frequently experienced them myself, although I had not expected to, I do think they are real illusions. Some are seen with high confidence. Also several subjects reported seeing coloured digits—even though they were not asked to report the colour of the digits and in fact had been told that they would always be black. Some binding errors may arise in memory; for example, a subject might remember seeing some red but forget where it was. But we still get a substantial number of illusory conjunctions when the task is simply to report whether there are two identical items in a display like that in figure 6.3b, where you can recombine the colour of the red H with the shape of the blue O to create an illusory red O, matching the one that is actually present.

Can we tie these binding errors more closely to the fact that we prevented subjects from focusing attention? In the experiment I described, where the relevant item was cued only after the display, subjects had no time to focus down to each coloured shape in turn in the brief exposures we gave them (around 200 ms). Using the same displays with another group of subjects, we cued the relevant item 150 ms before the display, and allowed subjects to ignore the digits, so that they could focus attention on the target item. We matched the overall accuracy by using a briefer exposure. As predicted, the binding errors disappeared: there were about as many intrusions (10% compared with 12%), which means that all the errors could be accounted for as misperceptions of the target feature or guesses. So it does seem that spatial attention plays a role specifically in the binding process.

In other experiments we have recorded similar errors that recombine parts of shapes, like lines and Ss that recombine to form dollar signs in displays like those in figure 6.4a,b, even when the lines must be taken from apparently holistic perceptual units like triangles (figure 6.4c). We get illusory arrows from lines and angles (figure 6.4d), but not illusory triangles (figure 6.4e), at least not until we add some circles to the display (figure 6.4f; Treisman and Paterson 1984). The explanation we proposed here is that triangles have the extra visual feature, closure, that also has to be present in the display before an illusory triangle can be generated.

The theory predicts that illusory conjunctions are created on the basis of the flags that signal the presence of particular features. If this is the case, the number of binding errors should not be affected by the similarity of the objects on other attributes. This is what we found. Differences in the shape or size of the objects made little difference to the probability of a binding error involving colour, as though the features are detected independently of each other and then bound. In generating the resulting percept, the spatial distribution of colour is selected to fit the shape with which it has been bound, whether correctly or erroneously.

Some researchers (see, for example, Cohen and Ivry 1989) have shown spatial proximity effects on illusory conjunctions, such that features are more likely to be wrongly bound if they are close in space than if they are distant. Cohen and Ivry suggested that features have 'coarse' location tags that are preattentively available. Proximity effects on binding errors could be a problem if we assume that locations are not available within the feature maps. However, it is very difficult to distinguish coarse coding of location from the idea that attention can rapidly zoom in to define a general area (like the upper left quadrant), but that it takes longer to focus more finely on one of two adjacent items. When the task prevented this zooming in by focusing attention narrowly at the fovea, our normal subjects showed no more illusory conjunctions between adjacent than between distant items (Treisman and Schmidt 1982). The Balint's patient, R.M., showed no effect
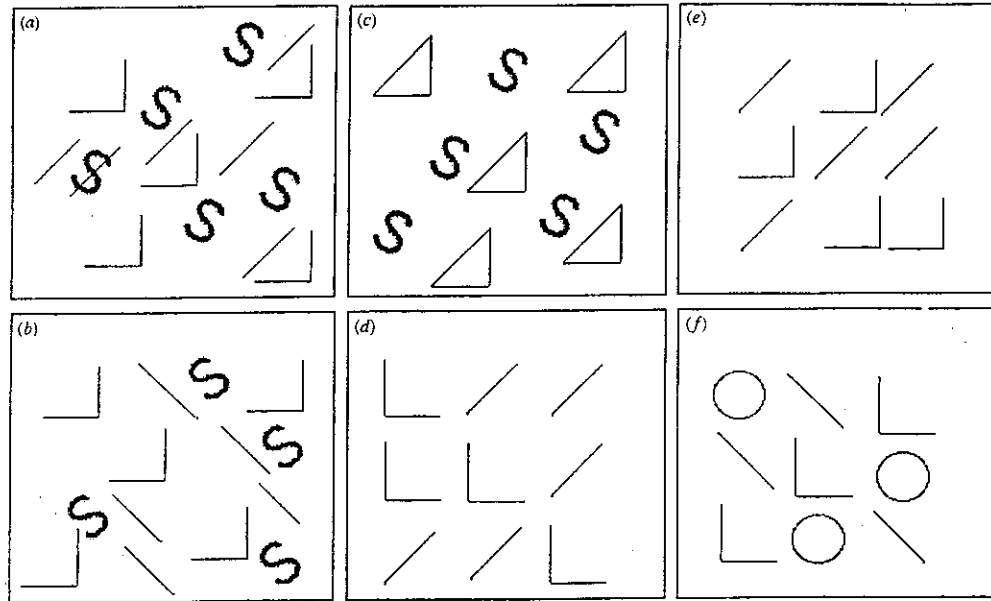
Fig. 6.4 Examples of displays used to look for illusory conjunctions of parts of shapes. Subjects report illusory dollar signs in 6.3a, illusory arrows in 6.3b, and illusory triangles in 6.3d but not 6.3c.

of distance on his binding error rates, consistent with the suggestion (§6.5) that he had lost his map of locations.

Another prediction is that the number of binding errors should also be independent of the number of instances of particular features because the claim is that all we have before the binding has occurred is information about the presence of features (the 'flags' in figure 6.1), not their individual instantiations. In an experiment (A. Treisman, unpublished data), we varied the number of instances of particular features in a display of four bars varying in orientation, format (filled, outline or dotted) and colour. To minimize memory errors, we cued subjects immediately after the display whether to report the digits (on 20% of trials) or one of the bars (on 80% of trials), giving high priority to accuracy in reporting the digits whenever they were cued. In displays with three instances of one feature (e.g. red) and only one of another (e.g. green), we found little difference in the number of illusory conjunctions involving migrations of the feature with three instantiations and of those with only one. The ratio was 1.5 : 1 rather than the 3 : 1 ratio that would be predicted if individuated tokens of the features were migrating. For example, in figure 6.5, reports of an illusory red bar in panels (a) and (b) were made on 15% and 10% of trials respectively, although there are three times as many red objects in panel (a). In another experiment varying just colour and orientation the ratio was even lower, 1.2 : 1. Note that to the extent that the amount of red present affects the chance of detecting it, quite apart from the number of instances of red, we would expect the ratio to exceed 1 : 1.

The evidence from illusory conjunctions supports four claims: (i) that features are separately coded, otherwise they could not recombine; (ii) that the binding problem is therefore a real one; (iii) that focused attention is involved in solving it; and (iv) that attention is not required for the
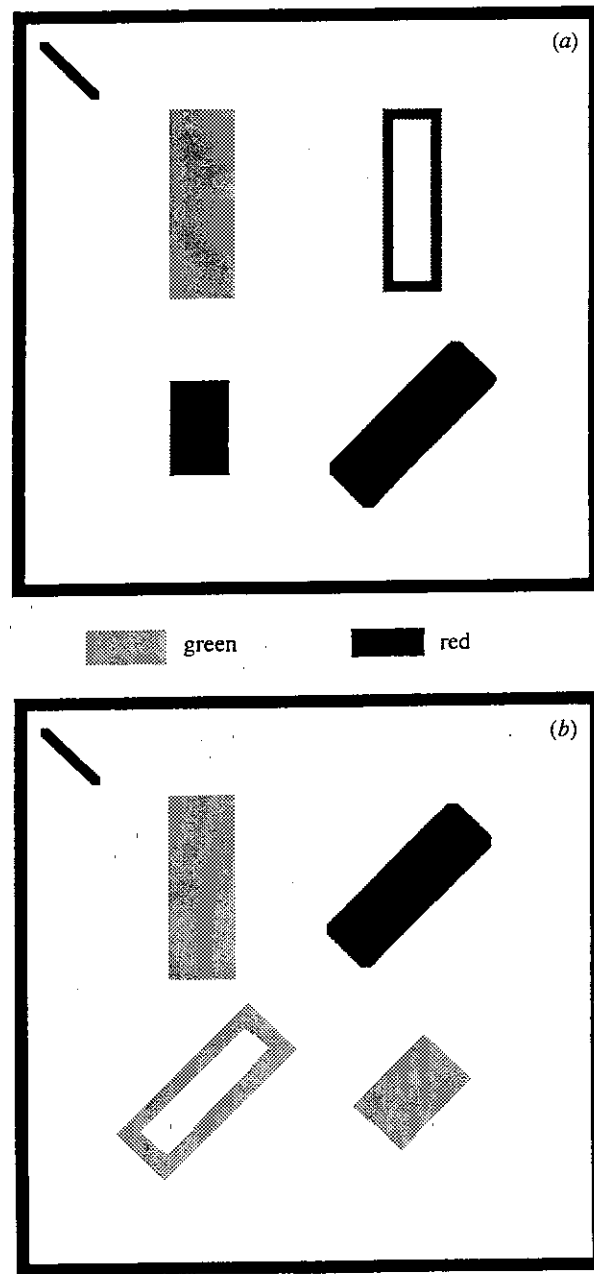
Fig. 6.5 Examples of displays to test whether feature tokens or feature types migrate when binding errors are formed. In reporting the colour of the cued bar, the token hypothesis predicts three times as many illusory migrations of red in *a* as in *b*.

simple detection of separate features (although it is often attracted by the prior detection of a unique feature).

6.4   Visual search and binding

Search tasks offer another source of information on the role of attention in binding. We can define a target so that it either does or does not require binding. In displays of green Xs and brown Ts, a target specified only by a conjunction of features, for example a green T, should require focused attention to each item in turn, whereas a target specified by either of two unique features, a blue letter or a curved letter, should not involve binding and might therefore be detected independently of attention. If attention must be focused on each item in turn to find conjunction targets, we predict a linear increase in search times with the number of items in the display. This is what we found (Treisman and Gelade 1980). On the other hand, feature targets, signalled by flags on the feature maps, showed no effect of the number of items. They simply popped out of the displays.

In some conjunction search tasks, there are other strategies besides the serial scan with focused attention, that can be used. We and others (see, for example, Treisman 1988; Wolfe *et al.* 1989; Nakayama 1990; Treisman and Sato 1990) have shown that when the target features are known in advance and when the relevant features are highly discriminable, subjects can use a feature-based grouping strategy to bypass the binding process. Essentially in looking for a red O among red Xs and blue Os, they may inhibit any location that contains blue and any location that contains an X. The red O would then emerge unscathed, without any need to bind distractor features. In the model, this would be implemented by reverse connections between the feature maps and the location map, selectively inhibiting all locations that contain unwanted features, and leaving only the target location to be checked (see figure 6.6).

6.5   Evidence from parietal lesions

So far this paper has been mostly behavioural. The model was developed from perceptual experiments. Can we tie it more closely to the brain? A classic experiment by Moran and Desimone (1985) seems consistent with the idea that attention selects by narrowing an attention window around the relevant object. Recording from cells in monkeys' area V4, they showed receptive fields essentially shrinking to exclude an unattended object when it fell into the same receptive field as the object to which the animal was trained to attend. What areas might control the attentional window? The parietal lobes are certainly involved in spatial attention. Unilateral damage to one parietal lobe produces a marked attentional deficit in the contralateral area of space. Particularly with right parietal damage, patients often show severe neglect of the left side of the visual field or the left side of an object. Investigations involving the patient I described earlier, who had such severe binding problems, may lead to a greater understanding of the brain areas involved. R.M. was unfortunate enough to suffer two successive strokes which destroyed both his parietal lobes, one after the other. He has normal acuity, contrast sensitivity, colour vision, stereopsis, and visual fields as tested formally by an ophthalmologist. However, he was left with a severe set of perceptual disabilities, some of which I've already outlined. Our hypothesis is that the master map of locations depends on parietal function. If this is correct, we can predict the set of deficits that the model would predict from his lesions.

1. Like any other theory, it would predict severe difficulties in conscious access to spatial information. The ability to point, reach for, or verbally label locations would be lost.
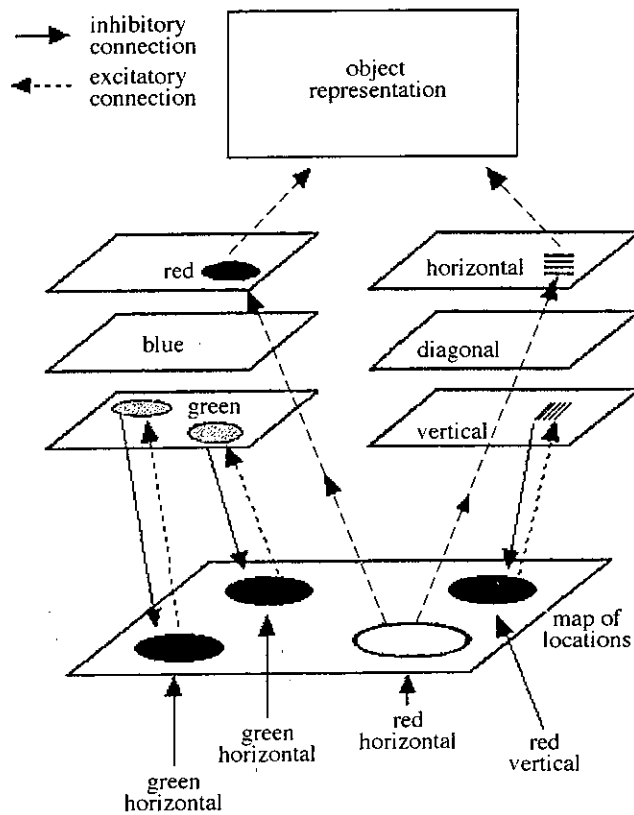
Fig. 6.6 Modified model to allow control of selective attention from one or more feature maps as well as through a serial scan by a spatial window.

2. If individuating objects depends on binding them to separate locations, only one should be visible at a time.

3. If space is the medium in which attention binds features, there is a risk of illusory conjunctions whenever more than one object is present.

4. Conjunction search should be abnormally difficult or impossible; however, in feature search tasks, there should be no difficulty in detecting the presence of a target feature, even when it is embedded in several nontargets, as feature detection does not depend on binding.

I have already described data from R.M. confirming prediction 3—the illusory conjunctions. It is important to note that R.M. has no difficulty in binding the features of single objects presented sequentially. When we presented two coloured letters successively for 3 s each (instead of simultaneously for a total of 10 s), the estimate of binding errors was 0. So R.M. does not have a general deficit in binding features or a memory problem in remembering them; his difficulty is specifically in the ability to bind simultaneously presented features to the correct objects.

We tested R.M. on prediction 4, and found a dramatic difference in his ability to do the tasks. The target was a red X. In the feature search condition, the distractors differed either in shape (red Os) or in colour (blue Xs). In the conjunction search task, he looked for a red X among red Os and blue Xs. He had no difficulty with the feature targets. He detected the unique colour or shape

in almost every trial, and independently of the number of items in the display, but he was unable to do the conjunction search, even with displays of only three to five items. He took up to 5 s and had error rates of 25%.

What about the first two predictions? These are the classic symptoms that Balint described in 1909. The spatial difficulties and the inability to see more than one object (simultanagnosia) are usually assumed to be separate and unrelated deficits, except that they co-occur with parietal damage. However, feature integration theory suggests that the simultanagnosia may actually be caused by the loss of space, simply because attention uses space to bind features to objects. R.M. did show severe deficits in his ability to localize. We asked him to report whether an X was at the top, bottom, left or right of the screen, or whether it was to the right, left, above or below an O. He was at chance in judging the relative locations of a widely separated X and O on the computer screen, and only slightly better at saying whether a single X was at the top, bottom, left or right of the screen. He seemed to have lost his representation of space almost completely. The confusion was not with the meaning of the words: he performed almost perfectly when asked to localize a touch on his back as up or down, or left or right. His somatosensory spatial functions were intact, and his spatial difficulties were specific to the visual modality.

R.M. also conformed to the second prediction. His simultanagnosia was as striking as his localization failures, at least in the early days after his stroke. When we held up two objects, say, a comb and a pair of scissors, he typically saw only one of them. When we asked him to count dots in displays of one to five, in an early session he saw at most two dots, even when we presented five. In later sessions, he guessed higher numbers but was still very inaccurate. We think he switched to attending more globally, to the group of dots as a whole, but this made him unable to access any individual dot and count it. R.M.'s simultanagnosia makes his normal performance in feature search all the more surprising. At a time when he could typically see only one or two objects, he had no difficulty detecting a unique colour or shape, regardless of how many other items were present.

An intriguing incident throws more light on R.M.'s perceptual experience. One morning he told us he had found a good way to improve his vision. With the help of his granddaughter, he had made a tube through which he looked at whatever he wanted to see more clearly. For someone suffering from simultanagnosia, one would think that a tube restricting the angle of vision would be the last thing they needed. However, on reflection, it makes more sense. If the damage to his brain prevents the normal binding of features without preventing their detection at early levels of visual processing, the features of different objects should tend to coalesce into a single object, producing confusing mixtures of several features in the one object that is seen. R.M. did complain of such illusions. For example, he said 'When I first look at it, it looks blue and it changes real quick to red. I see both colours coming together . . . Sometimes one letter is going into the other one. I get a double identity. It kind of coincides.' His descriptions sound as though he has no perceptual space in which to separate the letters and bind colours to shapes. The tube he invented may have helped by restricting the early detection of features to those of a single object. Essentially, he constructed an external window of attention to supplement a defective internal window. The findings with R.M. are consistent with the predictions that follow if he has lost the location map that controls spatial attention.

Further support for a parietal role, both in shifting spatial attention and in binding, comes from two recent studies of PET activation by Corbetta, Petersen and others (Petersen *et al.* 1994; Corbetta *et al.* 1995; see also Corbetta, this volume). They compared activation in a conjunction search task, and in a task that required active shifting of attention between locations to track a target. They found similar activation in the superior parietal cortex in both tasks, consistent with the prediction that the binding process required in conjunction search does involve scanning

with spatial attention, and that the parietal area is involved in its control. Ashbridge and co-workers (1997) found that transcranial magnetic stimulation to the right parietal lobe slowed conjunction search but left feature search unaffected. Taken together, these data suggest that we need an explicit representation of space for accurate conscious binding of features to objects. Thus, the dorsal parietal pathway interacts with the ventral pathway in mediating the perception of simultaneously presented objects. There might be 're-entrant' connections from the parietal lobes, perhaps via the pulvinar, to selected locations in visual areas V1 or V2.

6.6   Binding in feature search

Having shown a possible link between the model and the brain, I will describe next some further behavioural findings with normal subjects that led us to elaborate the theory. One challenge to the feature integration account of search came from a suggestion that the pattern of linearly increasing search times with increasing number of distractors might result when targets are difficult to discriminate from distractors because they are similar to them, and when the distractors are sufficiently different from each other to prevent good grouping and segregation from the target (Duncan and Humphreys 1989). This could account for the difficulty of conjunction search, in which the target shares one or other feature with all the distractors while the two distractor types differ in both features from each other. It also predicts that search for feature targets could be equally difficult if they closely resemble the distractors, although no feature binding should be involved. One can certainly get steep and linear search functions with targets defined by a small unidimensional difference (see, for example, Treisman and Gormican 1988). However, the critical question is what counts as a feature for the visual system. The answer must be determined empirically. The challenge led to some further research which pointed to a new version of the binding problem that might explain these data.

When I drew the model in figure 6.1, I put in three feature maps per dimension, mainly because drawing 50 was beyond my artistic capabilities. But there is some evidence that the visual system does use coarse coding, representing different values on each dimension by ratios of activity in only a few separate populations of detectors (for example, vertical, diagonal and horizontal, for orientation, or red, green, blue and yellow for colour). Stimuli differing only slightly along a single dimension would not activate separate populations of feature detectors and would not be expected to pop out. They would pose a somewhat different binding problem—binding to location—so that the small differences in activation between areas containing only distractors and an area containing the target as well could be discriminated.

We have observed large asymmetries in search difficulty with many different pairs of stimuli, depending on which is the target and which the distractors (Treisman and Gormican 1988). A tilted line pops out among vertical lines, a curved line among straight ones, a circle with a gap among closed circles, an ellipse among circles, whereas the reverse pairings—a vertical line among tilted ones, a straight line among curves, a closed circle among circles with gaps, and a circle among ellipses—give search that is much slower and seems to require focused attention to each item in turn. The targets that pop out behave as though they have a unique perceptual feature, like a red dot among green dots, whereas the others do not. We also find a marked search asymmetry when we compare search for a shape with an added feature (like a Q among Os) and search for a shape that lacks the same feature (like an O in a display of Qs, Treisman and Souther (1985)). Note that detecting an O among Qs also requires binding. To find the one circle which lacks a tail, we must locate all the tails and bind them to the Qs. On the other hand, to find a Q among Os, we can simply check for the presence of a tail anywhere in the display. There is no need to bind the tail to

know that a Q is present. In this example, the same discrimination poses very different problems for the visual system, depending on whether or not the task requires binding.

Can we find an analogy in the case of the search asymmetries with apparently unidimensional stimuli like the targets defined by orientation or curvature? Here is where the coarse coding of features becomes relevant. A slightly tilted line might be coded by activity mainly in the detectors for vertical, with some additional activity in the diagonal detectors, just as a Q can be described as an O plus an extra tail. A slightly curved line could be coded as basically straight, plus some additional activity in detectors for curvature. The hypothesis is that the presence of this extra activity is detected without any need to bind it to the object, and this is enough to signal that the target is present. On the other hand, when the target is the one vertical line in a display in which all the lines but one are slightly tilted, both vertical and tilted detectors would be active everywhere except in the one location where the vertical target leaves the tilt detectors silent.

If this hypothesis is correct, we should be able also to prevent pop-out for a tilted target by turning its detection into a task that requires binding. The assumption is that a tilted target pops out among verticals because of the additional unique activity it evokes in the detectors for diagonals. If we now mix diagonal distractors with the vertical ones, activity in the vertical and diagonal detectors must be bound together to identify the target by its particular ratio of activation levels in the two detector populations. This changes the task into search for a conjunction target, and we should expect to switch from parallel pop-out to serial search with focused attention. Similarly a purple target will pop out among either blue or red distractors alone, but among a mixture of blue and red distractors it will require binding of activity in blue and red detectors that share the same location and should therefore depend on serial attentive scanning. I tested search for purple targets tilted 27° in displays of blue vertical bars and pink bars tilted 63°, and found that search indeed looked serial, even though the colour and orientation of the target were objectively unique in the display and easily discriminable from either type of distractor alone (Treisman 1991). Furthermore, when we briefly flashed displays with the same stimuli, subjects made a large number of false alarms, detecting illusory targets. They saw far more illusory conjunctions in these displays in which I suggest that within-dimension binding is required than in displays where the target, although equated for similarity, was defined by a colour and orientation that would be directly coded by standard feature detectors, for example a blue vertical bar among purple and turquoise bars tilted 27° left and right. Thus, coarse coding by ratios of firing in different populations of feature detectors does seem to create another kind of binding problem.

The features that are preattentively detected are probably not those of the retinal image. Enns and Rensink (1991) and He and Nakayama (1992), have shown rapid or parallel detection of simple three-dimensional (3D) features of surfaces and illumination. The shading that results from 3D shape can produce segregation of a group of convex objects among concave ones, assuming lighting from above (Ramachandran 1988). This segregation is much clearer than with supposedly simpler black and white patterns. Although these might seem like conjunctions of shape and texture or luminance, the critical question is whether they are directly sensed by separate populations of neural detectors. Lehky and Sejnowski (1988) showed that simple neural networks, when trained to respond to gradients of shape from shading, evolve hidden units (i.e. units intermediate between the input and the output units) that look very similar to the simple cells that Hubel and Wiesel (1968) identified in area V1 and that have normally been assumed to be bar or grating detectors. The features that are directly coded by the visual system may actually be features that characterize 3D surfaces. It seems harder to find plausible candidates for featurehood in the geometric line arrangements of the cubes whose 3D orientations define the target in some of Enns and Rensink's (1991) experiments. However, what they find are asymmetries of search rather than the flat search functions associated with pop-out. Search is much faster for some target–distractor combinations

Table 6.1 Visual search studies

Serial search (attention required)	Parallel search (automatic pop-out)
Conjunction targets (e.g. green T in green Hs and brown Ts)	Feature targets (e.g. blue or S in green Hs and brown Ts)
Os in Qs	Qs in Os
Vertical line in tilted lines	Tilted line in vertical lines
Straight line in curved lines	Curved line in straight lines
Parallel lines in converging lines	Converging lines in parallel lines
Circle with gap in closed circles	Closed circle in circles with gaps
Purple bar tilted 27° in blue 0° (vertical) bars and red bars tilted 63°	Blue 0° (vertical) bar in purple bars tilted 27° left and turquoise bars tilted 27° right

than for others but not usually parallel unless shading is also present. Table 6.1 summarizes the results I have described on search and binding.

6.7   Implicit processing of conjunctions

For the last part of the paper, we move to another line of research which opens up issues in three new directions: so far I have discussed binding under a fairly narrow definition: it has been measured by conscious reports rather than by implicit indices; it has been manifested in immediate perceptual tests rather than in memory; and finally the information has concerned only stimulus features and locations, not the binding of actions to perceived events. Yet these are also aspects of a more general binding problem. We need to retain bindings in memory after the objects disappear, and we need to bind appropriate responses to the objects we identify. The experiments I have been pursuing recently extend the research in these three directions by comparing implicit with explicit measures of visual memory and specifying the choice of which object requires a response. The results have surprised us, and seem to have important implications for a more general understanding of binding.

First the question whether any implicit binding can be revealed, which we are unable to access consciously. There is increasing evidence that explicit reports may not exhaust all the information available to the visual system. Understanding what makes some information accessible and some not is an intriguing challenge. In particular, it is important to determine whether binding imposes a real computational limitation, or whether it is just a problem for conscious representation?

First, we did get a few surprising results with the patient R.M., when we used indirect measures to probe for implicit information about locations. We presented the word 'UP' or 'DOWN' at the top or bottom of a vertical rectangle. In any given trial, the semantics of the word and its location could be consistent (e.g. the word UP in the upper location), or inconsistent (e.g. the word UP in the lower location). R.M. read the words rapidly and correctly (note that binding was not necessarily involved in this reading task because the two words to be discriminated differed in all their letters and also in length), but his response times were 142 ms slower when the word was in an inconsistent location. So the locations interfered with his reading at a time when he was at chance in voluntarily locating the words. We were also able to show unconscious priming of spatial selection by a cross-modal cue although R.M. was unable voluntarily to select the cued item. We presented two visual letters, one on each side of the screen, at the same time as tapping his right or left shoulder. When we asked him to name the letter on the side that we had tapped,

he was at chance, but when we simply asked him to name the first (or only) letter he saw, he reported significantly more from the tapped side. It seems that some implicit representation of space remains despite the loss of parietal function, perhaps in extrastriate areas of the ventral pathway, although this information is not consciously accessible. Finally, Egly and colleagues (1995) ran another experiment that also revealed implicit information about the spatial distribution of elements. Their displays consisted of a global letter made of local letters. When asked what he saw, R.M. never reported the global letter. He seemed to see only one of the local letters. Yet when asked to classify the local letter as one of two possible targets, he was significantly slower when the global letter was inconsistent with the local ones than when it was consistent. Again, he seemed to have implicit information—this time about the whole, even though he could only respond to a part.

DeSchepper and I have explored implicit visual processing in normal subjects as well. We found indirect evidence that fairly complex patterns can be registered, bound and stored implicitly without conscious attention. However, it is important to add that this seems to be true only for one unattended object. When more are added, the evidence for implicit binding disappears (Neumann and DeSchepper 1992). The measure we used is known as negative priming. Subjects are typically shown two objects and asked to respond to one of the two, selected by some simple distinguishing feature like its colour. So, for example, Tipper and Cranston (1985) asked subjects to name pictures of familiar objects selecting the red one in each overlapped pair. Their responses were slower when the unattended green object on one trial became the attended red one on the next, relative to when two new objects were shown. A plausible account was that subjects inhibited the green object to avoid naming it instead of the red. When it then became the relevant object, on the next trial, they had to overcome the inhibition. As a result, the response was slightly delayed. This negative priming implies: (i) that subjects formed and retained a memory trace of the picture, even when it was the unattended member of a pair; and (ii) that an action tag was bound to the memory, specifying whether it should be responded to or ignored.

DeSchepper and I wondered whether novel patterns would also produce negative priming. If so, this would be evidence that a representation of their shape was formed, even in the absence of attention. We used overlapped pairs of 270 nonsense shapes that the subjects had never seen before, similar to those in figure 6.7 (DeSchepper and Treisman 1996). The task was to decide whether the green shape on the left matched a single white shape to the right of the display, ignoring the red shape. We gave subjects some practice with a set of 12 shapes that we used repeatedly, and then we introduced new shapes mixed with the old. We found a clear negative priming effect of about 30 ms, which was actually larger on the first trial in which a new shape was presented than it was for the familiarized shapes. Subjects apparently set up a new representation for the unattended as well as the attended shape and attached an inhibitory action tag to it, specifying that it should be rejected. On the other hand, they had no conscious memory at all for the unattended shape. Their recognition was at chance, even when probed immediately after a pair was presented. We infer that attention is needed at the time of encoding if objects are to be explicitly retrieved.

Can we say any more about the nature of these implicit memory traces? How detailed and specific are they? How well bound are their parts and properties? With two other students, Alex Kulczycki and Xuexin Zhang, I asked what happens if we change either a feature or a component of the unattended prime before presenting it again as the attended probe (Zhang *et al.* 1996). In one experiment, we either kept the size the same, or changed it to larger or smaller. Surprisingly, when we changed the size, the result was not inhibition but facilitation. Subjects were slightly quicker to respond to the previously unattended shape when its size was changed than to respond to a new shape. In another experiment, we presented only half of the prime shape combined with half a new shape, to see if the inhibition was attached only to the whole or separately to each component
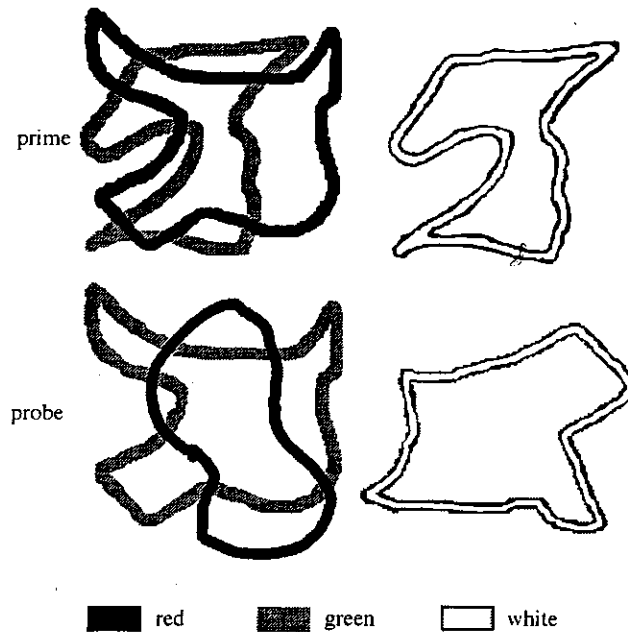
Fig. 6.7 Examples of a prime and a probe trial testing negative priming with novel nonsense shapes. Subjects attended to the green shape in each trial and decided whether it matched the white one. They ignored the red shape, which on negative priming trials, reappeared as the attended shape in green, resulting in slower responses.

part. Figure 6.8 shows three successive trials in which subjects matched the two outside shapes and ignored the central shape. Two trials later, half of the first ignored central shape appeared as half of the relevant outside shapes, combined with half of the second ignored shape. We got negative priming as usual when the shapes were identical, but again there was facilitation when we recombined two previously unattended halves from two different trials to form a new whole, and also when we presented only half the prime shape combined with a new half shape. Khurana and co-workers (1994) looked at negative priming for faces and got a similar result: when the prime and the probe face were identical, they got inhibition, but when they reversed the contrast on the probe trial, they got facilitation.

Thus, the action tag that produces negative priming seems to be bound to a very specific conjunction of shape with size or contrast and of the parts of a shape with each other. But in addition we form a more abstracted representation which is size and contrast-invariant, and which may have separate articulated parts. This representation is not linked to the specific responses required in the context in which it was seen, and it can facilitate later perception of similar or related objects.

We wondered how long these implicit memory traces for novel unattended shapes would linger in our subjects' brains. So we looked for negative priming, not only in the next trial, but after 10, 100 or 200 intervening trials with up to 400 different intervening shapes. To our considerable surprise, we found that the inhibition was undiminished 200 trials later. The binding here seems quite persistent even though it is formed in a single unattended trial. We also tested intervals of a day, a week and a month and found significant priming even at those long delays, but with some indication of a shift from inhibition to facilitation. The survival of these memory traces for novel shapes in our experiments suggests a surprising combination of visual plasticity and
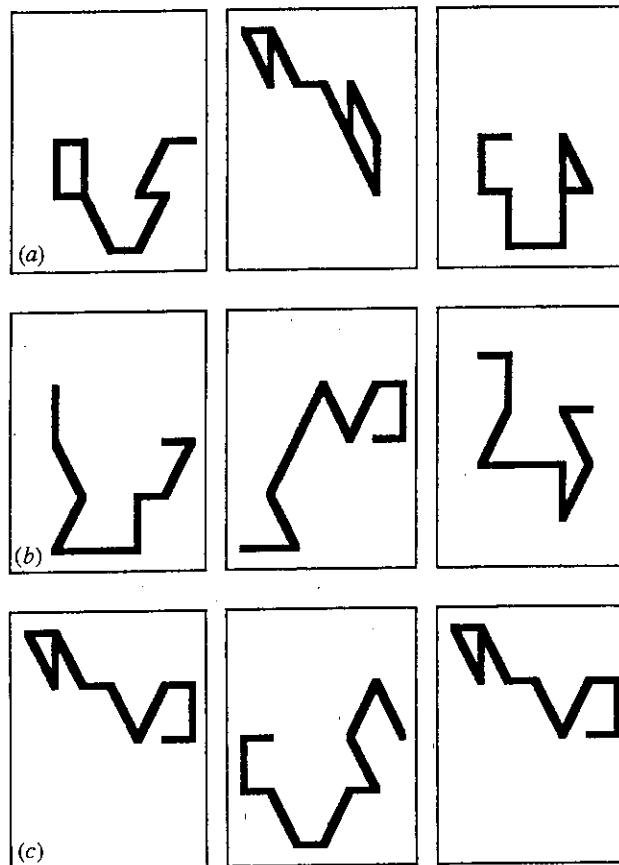
Fig. 6.8 Examples of two prime trials and a probe trial. The task was to decide as quickly as possible whether the shapes in the two outside positions were the same or different. The third (probe) trial recombines two previously unattended half primes to form the attended shape. Instead of negative priming, facilitation was observed relative to a new shape that had not been previously presented.

persistence. The newly formed representations remain available, separate and distinct from each other, for days or even weeks. If we had to speculate about where the representations are formed and stored, the temporal lobes in the ventral pathway seem a likely neural basis. Single unit recordings in monkeys, and brain imaging studies in humans, suggest that this is the area where objects are perceptually coded and form the memory traces that mediate subsequent priming. For explicit memory, on the other hand, the hippocampus and adjacent cortical areas are likely to be involved.

6.8 Conclusions

It is time to try to pull things together, both the different results described here, and their relation to the different view of attention proposed by Duncan (this volume). The results on tasks requiring explicit binding, both from normal subjects and from the patient R.M., suggested that attention is needed to bind features together, and that without attention, the only information recorded is the

presence of separate parts and properties. Yet both the negative priming results in normals and the interference from global letters in R.M. suggest that there are conditions in which wholes are registered automatically, without attention or conscious awareness.

I think there is no real contradiction between these results. Binding failures typically occur with high load displays when several objects must be processed under high time pressure. When there is only one unattended object, its features must belong together, so there should be no problem determining what goes with what. Indeed, when the number of unattended objects in a negative priming task was increased from one to two or three, Neumann and DeSchepper (1992) found that negative priming disappears. Attention here plays a different role: the unattended object may be suppressed to prevent response conflict and to ensure that conscious experience is coherent and consistent with our behavioural responses.

The patient R.M. makes binding errors with only two objects present because he is unable to separate them spatially to select one for attention and ignore the other. His implicit knowledge about locations that he cannot consciously locate may come from the ventral pathway where it is not normally accessible for conscious perception. As Humphreys (this volume) suggests, spatial relations within objects may be coded by separate systems from the spatial relations between objects. The arrangement of parts within an object may be coded holistically in order to identify the object rather than to relate the two parts as separate entities in space. We might explain the spatial Stroop interference shown by R.M., if his ventral pathway coded the rectangle together with its word as a consistent or inconsistent whole. In the location naming task, on the other hand, R.M. was asked to relate the position of the word as an object to the rectangle as another object, presumably requiring the use of his damaged parietal lobes to label these between-object relations.

The hypothesis outlined in this chapter about the role of attention in binding appears to conflict with Duncan's view of attention as integrated competition between objects (Desimone and Duncan 1995; Duncan, this volume; see also Bundesen, this volume). In Duncan's language, attention is a state into which the system settles through global competition between objects for dominance over experience and action, rather than a selective process that helps create those objects. He denies the need for any external selective mechanism. However, for the features of the same object to cooperate in competition with others and to benefit from each other's ascendancy in the competition, it seems as though they would already need to be bound. Duncan's suggestion is that binding is achieved through detectors directly coding conjunctions of pairs of attributes. An alternative is that the competition postulated in Duncan's framework could arise at a later level than the binding mechanism proposed here, and could have evolved to determine selection for the control of actions. Attention need not be a unitary process simply because a single word is used in everyday language.

One of the findings supporting Duncan's account is the long duration that he and his colleagues observed for the 'attentional blink' (the interference with detecting a second target caused by detecting the first in a rapid visual sequence of items (Duncan *et al.* 1994)). The interference lasts for about 500 ms after the first target has been detected, suggesting an 'attentional dwell time' of half a second rather than the 60 or so milliseconds implied when one interprets the slope of search times against the number of items simultaneously presented as a serial scanning rate. However, there are a number of ways of reconciling these findings. One is that scanning items that are simultaneously present and unmasked may allow some parallel preprocessing to occur at the same time as the serial attentional binding, whereas each item in a sequential display interrupts the processing of its predecessor and must itself be processed from scratch. With a presentation rate of 150 ms instead of 90 ms per item, we found little evidence of a blink, suggesting that when the processing is not interrupted by the early onset of a new item or mask, an upper limit

to the dwell time is 100–200 ms rather than 500 ms. Another contrast with most attentional blink experiments is that processing in search displays is serial across distractors, not targets, if it is serial at all. Processing of targets is likely to take longer than processing of distractors, as it requires commitment to a response. Finally, if processing in search is serial across pairs or small groups of homogeneous distractors rather than single items, observed slopes of 60 ms per item would imply dwell times of 120 or 180 ms (Treisman and Sato 1990), reducing the apparent discrepancy between sequential and simultaneous presentation.

As must be clear, this is work in progress and there is much that is not yet understood. The use of implicit priming measures opens new perspectives. The intriguing dissociations that we and others are finding between conscious experience and indirect priming suggest that the binding problem may be intimately bound up with the nature of consciousness, but that is a story that I think no one is yet ready to tell.

Acknowledgements

References

Ashbridge, E., Walsh, V. and Cowey, A. 1997 Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia* 35, 1121–31.

Cohen, A. and Ivry, R. 1989 Illusory conjunctions inside and outside the focus of attention. *J. Exp. Psychol. Hum. Percept. Perf.* 15, 650–63.

Corbetta, M., Miezin, F., Dobmeyer, S., Shulman, G. and Petersen, S. 1991 Selective and divided attention during visual discrimination of shape, colour and speed: functional anatomy by positron emission tomography. *J. Neurosci.* 11, 2383–402.

Corbetta, M., Shulman, G. L., Miezin, F. M. and Petersen, S. E. 1995 Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science* 270, 802–5.

Cowey, A. 1985 Aspects of cortical organization related to selective attention and selective impairments of visual attention. In *Attention and performance XI* (eds M. P. and O. Marin), pp. 41–62. Hillsdale, NJ: Erlbaum.

Damasio, A., Yamata, T., Damasio, H., Corbetta, J. and McKee, J. 1980 Central achromatopsia: behavioral, anatomic, and physiologic aspects. *Neurology* 30, 1064–71.

DeSchepper, B. and Treisman, A. 1996 Visual memory for novel shapes: implicit coding without attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 27–47.

Desimone, R. and Duncan, J. 1995 Neural mechanisms of selective visual attention. *A. Rev. Neurosci.* 18, 193–222.

Duncan, J. and Humphreys, G. W. 1989 Visual search and stimulus similarity. *Psychol. Rev.* 96, 433–58.

Duncan, J., Ward, R. and Shapiro, K. 1994 Direct measurement of attentional dwell time. *Nature* 369, 313–15.

Egly, R., Robertson, L. C., Rafal, R. and Grabowecky, M. 1995 Implicit processing of unreportable objects in Balint's syndrome. Los Angeles: Psychonomic Society Abstracts.

Enns, J. and Rensink, R. A. 1991 Preattentive recovery of three-dimensional orientation from line drawings. *Psychol. Rev.* 98, 335–51.

Friedman-Hill, S. R., Robertson, L. C. and Treisman, A. 1995 Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science* 269, 853–5.

Goodale, M. A. and Milner, A. D. 1992 Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–5.

Gray, C. M., Konig, P., Engel, A. and Singer, W. 1989 Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–7.

Gulyas, B. and Roland, P. E. 1991 Cortical fields participating in form and colour discrimination in the human brain. *NeuroReport* **2**, 585–8.

He, Z. J. and Nakayama, K. 1992 Surfaces versus features in visual search. *Nature* **359**, 231–3.

Hubel, D. H. and Wiesel, T. N. 1968 Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**, 215–43.

Hummel, J. E. and Biederman, I. 1992 Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* **99**, 480–517.

Khurana, B., Smith, W. C., Baker, M. T. and Huang, C. 1994 Face representation under conditions of inattention. *Invest. Ophthalmol. Vis. Sci.* **35**(Suppl. 4), Abstract No. 4135.

Lehky, S. R. and Sejnowski, T. J. 1988 Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature* **332**, 154–5.

Meadows, J. C. 1974 Disturbed perception of colours associated with localized cerebral lesions. *Brain* **97**, 615–32.

Moran, J. and Desimone, R. 1985 Selective attention gates visual processing in the extra-striate cortex. *Science* **229**, 782–4.

Nakayama, K. 1990 The iconic bottleneck and the tenuous link between early visual processing and perception. In *Vision: coding and efficiency* (ed. C. Blakemore), pp. 411–22. Cambridge University Press.

Navon, D. 1977 Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* **9**, 353–83.

Neumann, E. and DeSchepper, B. G. 1992 An inhibition-based fan effect: evidence for an active suppression mechanism in selective attention. *Can. J. Psychol.* **46**, 1–40.

Petersen, S. E., Corbetta, M., Miezin, F. M. and Shulman, G. L. 1994 PET studies of parietal involvement in spatial attention: comparison of different task types. *Can. J. Exp. Psychol.* **48**, 319–38.

Posner, M. I. 1980 Orienting of attention. *Q. J. Exp. Psychol.* **32**, 3–26.

Ramachandran, V. S. 1988 Perceiving shape from shading. *Sci. Am.* **259**, 76–83.

Robertson, L., Treisman, A., Friedman-Hill, S. and Grabowecky, M. 1997 The interaction of spatial and object pathways: evidence from Balint's syndrome. *J. Cogn. Neurosci.* **9**, 254–76.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J., Brady, T. J., Rosen, B. R. and Tootell, R. B. H. 1995 Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–93.

Singer, W. and Gray, C. M. 1995 Visual feature integration and the temporal correlation hypothesis. *A. Rev. Neurosci.* **18**, 555–86.

Tanaka, K. 1993 Neuronal mechanisms of object recognition. *Science* **262**, 685–8.

Tipper, S. P. and Cranston, M. 1985 Selective attention and priming: inhibitory and facilitatory effects of ignored primes. *Q. J. Exp. Psychol.* A **37**, 591–611.

Treisman, A. 1982 Perceptual grouping and attention in visual search for features and for objects. *J. Exp. Psychol. Hum. Percept. Perf.* **8**, 194–214.

Treisman, A. 1988 Features and objects: the Fourteenth Bartlett Memorial Lecture. *Q. J. Exp. Psychol.* A **40**, 201–37.

Treisman, A. and Gelade, G. 1980 A feature integration theory of attention. *Cogn. Psychol.* **12**, 97–136.

Treisman, A. and Gormican, S. 1988 Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.* **95**, 15–48.

Treisman, A. and Paterson, R. 1984 Emergent features, attention and object perception. *J. Exp. Psychol. Hum. Percept. Perf.* **10**, 12–21.

Treisman, A. and Sato, S. 1990 Conjunction search revisited. *J. Exp. Psychol. Hum. Percept. Perf.* **16**, 459–78.

Treisman, A. and Schmidt, H. 1982 Illusory conjunctions in the perception of objects. *Cogn. Psychol.* **14**, 107–41.

Treisman, A. and Souther, J. 1985 Search asymmetry: a diagnostic for preattentive processing of separable features. *J. Exp. Psychol. Gen.* **114**, 285–310.

Ungerleider, L. G. and Mishkin, M. 1982 *Two cortical visual systems*. Cambridge, MA: MIT Press.

Ward, L. 1982 Determinants of attention to local and global features of visual forms. *J. Exp. Psychol. Hum. Percept. Perf.* **8**, 562–81.

Wolfe, J. M. 1998 Inattentional amnesia. In *Fleeting memories* (ed. V. Coltheart). Cambridge, MA: MIT Press.

Wolfe, J. M., Cave, K. R. and Franzel, S. L. 1989 Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perf.* **15**, 419–33.

Zeki, S. M. 1991 Cerebral akinetopsia (visual motion blindness). *Brain* **114**, 811–24.

Zeki, S. 1993 *A vision of the brain*. Oxford: Blackwell.

Zihl, J., von Cramon, D. and Mai, N. 1983 Selective disturbance of movement vision after bilateral brain damage. *Brain* **106**, 313–40.