Perception, 2001, volume 30, pages 1227 – 1248

DOI:10.1068/p3133

Can attention select only a fixed number of objects at a time?

Greg Davis, Victoria L Welch, Amanda Holmes, Alex Shepherd
Centre for Brain and Cognitive Development, Department of Psychology, Birkbeck College, Malet Street, London WC1E 7HX, UK; e-mail: g.davis@psyc.bbk.ac.uk
Received 18 September 2000, in revised form 7 June 2001

Abstract. Several previous studies have suggested that we may attend only a fixed number of 'objects' at a time. However, whereas findings from two-target experiments suggest that we can attend only one object at a time, other results from object-tracking and enumeration paradigms point instead to a four-object limit. Here, we note that in these previous studies the number of objects covaried with the overall size and complexity of the stimulus, such that apparent one-object or four-object limits in those tasks may reflect changes in the complexity of attended stimuli, rather than the number of objects per se. Accordingly, in the current experiments we employ stimuli in which the number of objects varies, while overall size and complexity are held constant. Using these refined measures of object-based effects, we find no evidence for a one-object or four-object limit on attention. Indeed, we conclude that the number of attended objects does not affect how efficiently we can attend a given stimulus. We propose and test an alternative approach to object-based attention limitations based on within-object and between-object feature-binding mechanisms in human vision.

1 Introduction

Natural scenes can often comprise dense arrays of colours, textures, and shapes, belonging to a variety of different objects. In order to guide our behaviour efficiently in such complex environments, human vision must be able to select visual information that is currently relevant to us, while ignoring other irrelevant information. To achieve this, the human visual system incorporates a variety of different selective mechanisms, collectively termed 'attention', which enhance processing of particularly relevant objects (eg Duncan et al 1997; Moran and Desimone 1985).

In several earlier studies, attention was assumed to behave like a 'spotlight' or 'zoom lens', enhancing processing within an approximately circular region of the visual field independently of the objects situated there (eg Eriksen and Eriksen 1974). However, in natural scenes, where objects can be densely packed together and can overlap one another, a spotlight may not be an optimal strategy for selection. To illustrate, suppose that the vertically orientated bar in figure 1a were to be relevant to our current goals, and the horizontally oriented bar, irrelevant. In such a case, if attention were to behave like a circular spotlight, then the minimum circular spotlight required to select the relevant vertical bar would be that indicated by the dotted outline in the figure. Note that such a spotlight could not select the relevant vertical bar without also selecting the irrelevant horizontal bar. Thus, a spotlight would not adequately select relevant information from irrelevant information.

In response to the potential limitations of an attention spotlight, several authors have proposed that attention might be 'object-based' (eg Duncan 1984; Humphreys 1998; Pylyshyn 1989). On such a view, early stages of visual processing 'bind' together multiple features (aspects of colour, texture, edges, etc) to form unitary representations conventionally referred to as 'objects' in the literature (see eg Duncan 1984). In the case of figure 1a, all the visual features pertaining to the vertical bar (long vertical edges, dark grey colour, etc) are bound together to form one object, and all the features of the horizontal bar (long horizontal edges, dark grey colour, etc) form another object.

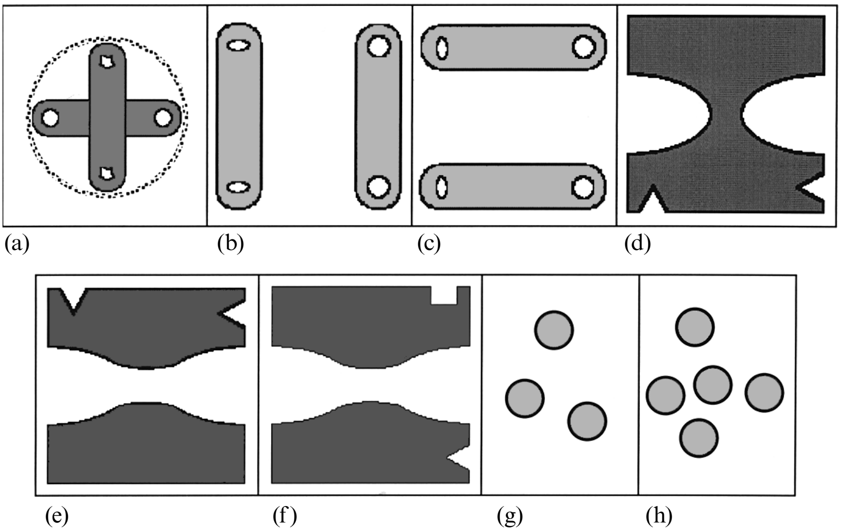

Figure 1. (a) A circular attentional 'spotlight', illustrated by the dotted-outline circle, cannot select the vertical bar without also selecting the horizontal bar (see text). (b) and (c) Typical two-target displays comprising two objects. In (b), both target features (the squashed-ring elements, top left and bottom left) appear on a single object, whereas in (c), they appear on two separate objects. (d) A single-large-object display from Davis et al (2000). (e) and (f) Two-small-object displays from Davis et al (2000): In (e), both target features (triangular notches, top left and top right) appear on a single object, whereas in (f), the notches (square notch top right, triangular notch, bottom right) appear on two separate objects. (g) and (h) Typical displays from object-enumeration studies, comprising three and five objects, respectively.

If attention were able to select one of these objects at a time (rather than a circular region of space) then it could efficiently select the relevant vertical bar (in our imagined situation relating to figure 1a) without also selecting the irrelevant horizontal bar. For our attention to select objects rather than circular regions of space therefore seems to make functional sense as it would allow us to attend relevant objects efficiently while ignoring irrelevant objects, even when they overlap each other.

Numerous previous studies, some of which are described below, have provided evidence that attention tends to select objects rather than circular regions of space (eg Duncan 1984; Egly et al 1994; Lavie and Driver 1996; Watson and Kramer 1999). However, while it is now widely agreed that at least some aspects of attention are object-based, there is less agreement regarding the specific effects of objects upon attention. Here, we address an enduring controversy in attention research concerning whether the amount of information we can attend at once is limited to a fixed, 'magical' number of objects (eg Duncan et al 1997; Sears and Pylyshyn 2000). Two distinct types of behavioural evidence have been interpreted as evidence in favour of such a view: two-target studies and tracking and enumeration paradigms. While both of these provide apparent evidence that our attention selects a specific number of objects at a time, and not a circular region of space as suggested by the spotlight metaphor, they provide conflicting indications as to how many objects we can attend. As we discuss below, findings from two-target studies suggest that we can efficiently attend only one object at a time, such that significant costs on performance arise when two objects must be attended concurrently. In contrast, tracking and enumeration studies appear to indicate that up to four objects can be attended simultaneously with little or no performance cost compared with attending only one or two objects.

1.1 *Two-target studies*

In 'two-target' studies, observers make speeded (or near-threshold accuracy) judgments regarding pairs of target features (eg shape, colour, or texture elements). Typically, when both target features can be considered to belong to the same object they are more efficiently judged than when they belong to different objects, even when other spatial factors are strictly controlled for (eg Behrmann et al 1998; Duncan 1984; Lavie and Driver 1996; Watson and Kramer 1999). To illustrate, consider a two-target task in which observers must make same – different judgments concerning the two squashed-ring elements in figures 1b and 1c (one at the top left of each figure, one at the bottom left). Although the two target features (the two squashed-ring elements) are equidistant from each other in the two displays, they would typically be judged more efficiently in figure 1b, where they both belong to one object, than in figure 1c, where they belong to two separate objects. Spotlight views of attention are unable to explain this pattern of performance, as a circular spotlight could encompass the features equally well in each display. Instead, the performance benefits for judging features from the same object versus different objects have been interpreted by some authors as evidence that we can efficiently attend to the features of only one object at a time (eg Duncan 1984; Duncan et al 1997; Humphreys 1998). When the two target features appear on two separate objects (as in figure 1c), both of the two objects must always be attended (either simultaneously or consecutively) in order to select the target features. However, when both target features appear on the same object (as in figure 1b), observers may often be able to restrict their attention to just that one object, ignoring the other object in the display. If attending to two objects is inherently more difficult somehow than attending only one object, then this would (correctly) predict better performance when features appear on one object than when they appear on two objects.

1.2 *Enumeration and tracking tasks*

A second source of evidence for object-based attention derives from tasks where observers must enumerate or attentionally 'track' a small number of items in a given display. In some studies, irrelevant distractor items are interspersed between the task-relevant items such that if attention were to behave as a spotlight, it could not select the relevant items without also selecting the irrelevant distractors. On a spotlight view of attention, therefore, the introduction of distractors should greatly affect performance. However, such is not the case; our ability to track and enumerate multiple items is evident even when irrelevant distractor items are present, suggesting that attention is able to select the relevant items in an object-based manner (see eg Pylyshyn and Storm 1988; Scholl 2000). Enumeration and tracking studies therefore provide evidence for object-based views of attention over spotlight-based approaches, in accord with conclusions based on two-target studies.

However, whereas two-target findings are easily accounted for by assuming that we can attend only one object at a time, enumeration and tracking results are not. Generally, observers can accurately enumerate up to four items with little cost compared with two or three items. Only when five or more items must be counted, does performance deteriorate significantly (eg Trick and Pylyshyn 1993). Similar patterns of performance are yielded in tasks where moving items prespecified as targets must be tracked by attention among other identical distractors: up to four can be tracked with relative ease, but tracking five items or more is significantly harder (eg Pylyshyn 1989; Scholl and Pylyshyn 1999). Such findings suggest that up to four separate 'items' can be attended efficiently at once. Although each 'item' may not correspond to one object, models which assume that we can attend four objects efficiently at a time seem better able to explain these results than models assuming we can attend only one.

1.3 *How many objects can we attend at once?*

There is thus an apparent incongruity between findings from two-target studies versus from enumeration and tracking paradigms. While two-target findings suggest that we can attend to only one object efficiently at a time, incurring significant performance costs when two objects must be attended concurrently, tracking and enumeration studies seem to indicate that up to four objects can be attended with relatively little cost compared to only one or two objects. One possible route to reconciling these two different findings comes from Davis et al (2000) who reappraised two-target evidence that we can efficiently attend only one object at a time. To illustrate their argument, consider the two objects in figure 1b, which illustrates a typical two-target-study display. As the two objects are of equal size, the two objects between them necessarily constitute roughly twice the amount of 'perceptual information' as comprised by one of the two objects. Thus, if, as recent evidence suggests, attention selects whole objects, then when two objects are selected approximately twice as much 'perceptual information' must be attended compared with when attention is focused upon just one of the objects.[(1)]

This simple observation renders many two-target findings ambiguous with regard to whether attending two objects is easier than attending one object. Any performance costs that arise in typical two-target studies when two objects must be attended versus only one may arise *either* because attending two objects is inherently more difficult than attending only one *or*, alternatively, because two objects in those displays comprise more information than one object. That is, the apparent difficulties in attending to two objects versus one in two-target studies might in fact reflect the greater amount of attended 'information' comprised by two objects and not the greater number of attended objects per se.

In response to this ambiguity, Davis and colleagues devised stimuli in which the number of objects could be manipulated, with other factors remaining approximately constant. Figures 1d and 1e illustrate versions of their stimuli. Figure 1d comprises a single large object, while figure 1e contains two smaller objects, but the two displays are otherwise very alike, being approximately equivalent in terms of overall surface area and amount of 'information'.[(2)] If attending two objects is inherently more difficult than attending only one then, as the two display types are equivalent in all other respects, the two-object displays should be attended less efficiently overall than the displays comprising one large object. However, if two objects can be attended as efficiently as one object, once other factors (such as amount of 'information') are held constant, the two displays should be equally well attended overall (as they comprise approximately the same amount of information).

In order to measure how effectively their one-object and two-object displays could be attended, Davis et al asked observers to make speeded same–different discriminations concerning pairs of notches on the objects (see figures 1d–1f), a task similar to that employed by many previous two-target studies. However, before any direct overall comparison of the one-object versus two-object displays could be attempted, it was necessary to ensure that observers really had perceived the two-object displays to comprise two separate objects, but the one-object displays to comprise only one object. Davis et al therefore examined performance for individual conditions within the

[(1)] Note that this objection also applies to the studies of Driver and Baylis (1993). They employed identical stimuli for the two-object versus one-object conditions, manipulating which regions of the stimulus observers must attend. However, even in that study, the two objects comprised more perceptual information spread over a larger area than did the one object.

[(2)] Note that it is the amount of *attended* information that is crucial here; this would not incorporate information from ignored distractor items. Additionally, while the concept of 'amount of information' is deliberately vague for our current purposes, we later propose a more detailed index of this concept that is more open to empirical test.

two-object displays where target features appeared on a single small object (figure 1e) versus on two small objects (figure 1f), and performance on similar conditions within the one-object displays.

1.3.1 *Two-object displays.* Davis et al reasoned that, if their two-object displays had been perceived to comprise separate objects, then patterns of performance within two-object displays should mirror those in previous two-target studies. First, when the two target features were horizontally separated and appeared on a single small object (as in figure 1e), observers should often be able to effectively confine their attention to just that object, thus attending to only half the information in the display. In many previous two-target studies, such conditions were found to yield efficient attention and enhanced performance. However, when the two features were vertically separated and appeared on two separate objects (as in figure 1f), observers would always have to attend to both objects, twice the amount of information as comprised by a single object, and performance should deteriorate accordingly.

Note that for this comparison of individual conditions within the two-object displays, attending to two objects required that twice as much information be attended compared with attending one object, as in previous two-target studies (as in figures 1b and 1c). With this confound present, any advantages for trials where features appeared on a single small object relative to when they appeared on two separate objects could either be due to the number of objects attended, or (as Davis et al contended) the amount of information attended. Thus, as for previous two-target studies, this comparison could not determine whether the number of attended objects per se, or alternatively the amount of information attended was responsible for the patterns of performance found there. Nonetheless, if the Davis et al two-object displays showed the classic pattern of performance for pairs of small objects (an advantage for features appearing on the same object), this would provide an indication that observers had perceived these displays to comprise two separate objects.

1.3.2 *One-object displays.* There was only a minor physical difference between the one-object and two-object displays employed by Davis et al, a small connecting region present only in the one-object displays (compare figures 1d and 1e). However, if this minor difference had caused observers to perceive the one-object displays as comprising a single large object, these displays should show markedly different patterns of performance to the two-object displays. Horizontally separated and vertically separated features in the one-object displays (see eg figure 1d) now always appeared on a single large object, such that in neither case could attention be confined to one half of the display or the other (as attention is compelled to select the whole object). For the one-object displays therefore, no differences in performance should be observed for horizontally versus vertically separated features, in contrast to the predicted performance for the two-object displays where an advantage for horizontally separated features (belonging to the same object) was predicted.

An absence of any performance differences between horizontally separated and vertically separated features in the one-object displays would also provide confirmation that any performance patterns observed in the two-object displays did not reflect differences in perception of horizontally separated versus vertically separated features per se, as such differences should apply equally to the one-object and two-object displays. Rather, any performance patterns observed in the two-object displays would have to be related to the perception of separate objects in those displays.

1.3.3 *Comparison of one-object versus two-object displays.* Davis et al found the expected patterns of performance in both one-object and two-object displays, lending support to the claim that the two-object displays had indeed been perceived as two small objects, and the one-object display as one large object. Note that both (i) theories which assume a one-object limit, and (ii) the Davis et al alternative account (where there is no one-object limit) are able to explain the Davis et al findings within each display. However, the two approaches make contrasting predictions when overall performance is compared between one-object and two-object displays. If attention can select only one object efficiently at a time, two-object displays should be less efficiently attended overall (ie including trials where features fell on the same object or on different objects in those displays) than one-large-object displays, even though the amount of information to be attended is now approximately equated in the two cases. In contrast, if attention can in principle select two objects as efficiently as one object, once the amount of attended information is equated in the two cases, equal overall performance should now be predicted for the one-object and two-object displays. In two separate studies, Davis et al found equivalent performance overall for one-object versus two-object displays, which suggests that two objects can in principle be attended as efficiently as one object, when the information to be attended is equated in the two cases.

1.4 *Can we attend four objects at a time?*

The Davis et al (2000) findings are incompatible with the claim that we can attend only one object efficiently at a time, but are compatible with the rival claim stemming from enumeration and tracking studies that up to four objects may be attended at a time without significant cost. These findings therefore provide one possible resolution to the issue of how many objects attention can efficiently select at once. However, many previous enumeration and tracking studies are subject to the same 'amount of perceptual information' criticisms as Davis and colleagues noted for two-target studies. Consider figures 1g and 1h, which comprise three and five objects, respectively. On any plausible index of perceptual information, the five objects in figure 1h must comprise more information overall than the three objects in figure 1g. Thus, any costs for attending five versus three objects in such displays may either result from the greater number of objects per se to be attended in the five-object display, or alternatively from the greater amount of information to be attended there (see footnote 2). Tracking and enumeration findings of this kind are thus somewhat ambiguous with regard to the presence or absence of a four-object limit on attention. The primary aim of our first study was to provide a new test of a potential four-object limit on our attention under conditions where the number of objects alone could be varied selectively, while the overall amount of information remained approximately constant.

In order to measure any potential four-object limit we adapted the Davis et al stimuli to yield two novel display types; these each comprised the same amount of perceptual information overall, but differed in the number of objects they contained. Specifically, our new displays were constructed from three smaller versions of the Davis et al one-object and two-object stimuli, yielding three and six objects, respectively (see figures 2a and 2b).

In addition to basing our stimuli on the patterns employed by Davis et al, we also employed a two-target task similar to that used in those studies. In many previous reports, two-target findings had suggested a one-object limit, while tracking and enumeration studies had suggested a four-object limit. However, the numerous differences between the various paradigms made this apparent contradiction hard to interpret. For example, in two-target paradigms, the task requires observers to make explicit judgments about the features (eg colour, depth, shape) of each attended object, whereas for tracking and enumeration tasks each object can be treated as an 'atom', with no requirement to explicitly access the object's individual features. It is possible therefore, that we can track several objects at a time, but can access the features of only one (see eg Humphreys 1998). Alternatively, the different size or complexity of typical displays in two-target versus enumeration and tracking paradigms may cause attention to operate differently in the two cases. Accordingly, to ensure that our test of a potential four-object limit would be more directly comparable with the Davis et al (2000) test of a one-object limit, we applied a very similar methodology in our current study to that employed by Davis and colleagues.

2 Experiment 1: attention to three versus six objects

In experiment 1 the presented objects were similar to those in the Davis et al (2000) study, but were reduced to approximately one-third of the size used by Davis et al. Additionally, displays now comprised either three large objects (as in figure 2a) or three pairs of small objects (ie six objects, as in figure 2b). That is, in each three-object display, three small versions of the one-large-object stimuli from Davis et al were presented, and in each six-object display three versions of the two-small-objects stimuli from Davis et al were presented. These stimuli could be arranged in a triangle (figures 2a – 2c) or an inverted triangle (figure 3) formation.

The six objects or three objects in a given trial were initially presented for 2.4 s without any notches removed (see figure 3, frame 1), before the target features were then presented. The possible target features were a small square notch on the corner of an object (see enlargement in figure 2d) or a small diagonal notch on the corner of an object (see enlargement in figure 2e). Two features were presented on every trial, and the task was again to determine whether the features were the same or different. The two features could appear on any two corners of any of the objects within the display, and the position of one feature had no predictive value as to the likely location of the other feature. The features were presented until response (8 observers) or for 224 ms (7 observers) to examine whether saccades from one feature to another played a role in our findings. For observers where features were presented until response, saccades to one feature then the other were possible. However, for observers where features were presented for just 224 ms, this duration was too brief for such saccades to operate effectively (in this case the display was replaced by a masking display in which all possible feature locations were masked: see figure 3, frame 3 for a six-object example). If the same patterns arose for both groups, we could ensure that saccades played no role in our findings.

As for the Davis et al study, we also intended to compare performance in individual conditions within the three-object and six-object displays to confirm that they were indeed treated by the visual system as comprising three and six objects, respectively. In the six-object displays, we compared response times (RTs) for two trial types. The first of these arose when the two target features were horizontally displaced from each other, and (by chance) appeared on the same small object (eg figure 2c). The second arose when the two features were vertically displaced and appeared on two separate objects, but within the same ‘pair’ of objects as each other (ie on neighbouring objects corresponding to a single version of the Davis et al two-object stimuli, eg figure 3). We expected that the former (within-object, horizontally displaced features) condition should yield faster RTs than the latter (between-object, vertically displaced features) condition, a typical pattern of results for pairs of perceptual objects [see the description of the Davis et al (2000) study in section 1]. However, as in the Davis et al study, this pattern should not hold for corresponding comparisons between horizontally versus vertically displaced features in the three-large-objects displays, since features in both of those conditions should pertain to the same perceptual object. Such a pattern of performance would indicate that the six-object displays had not been perceived to contain three large objects (or Gestalts), but rather comprised six individual objects (ie three pairs of small objects), whereas the apparently minor differences between the two display types had caused observers to perceive three large objects in the three-object displays.

Note that this comparison of features on one large object (in the three-object displays) versus within a 'pair' of objects (in the six-object displays) followed exactly the same logic as the comparison of one large object versus two small objects in Davis et al (2000).

If our attention is subject to a four-object limit then we should expect the three-object displays to be efficiently attended. However, since the six-object displays should comprise too many perceptual objects to be attended efficiently on this four-object account, they should now be attended far less efficiently.

2.1 *Method*

2.1.1 *Observers.* Fifteen new observers from the department subject panel were recruited. Eight were female, seven were male, and their ages ranged from 20 to 26 years, with a mean of 23 years. Each was paid five pounds.

2.1.2 *Displays.* The stimuli were presented on a Sony 17 inch screen with a Power Macintosh G3 computer running 'Vscope' experiment-generator software (Enns and Rensink 1992). Figure 2a illustrates a typical display in the three-large-objects condition, while figures 2b and 2c illustrate typical six-small-objects displays. These figures

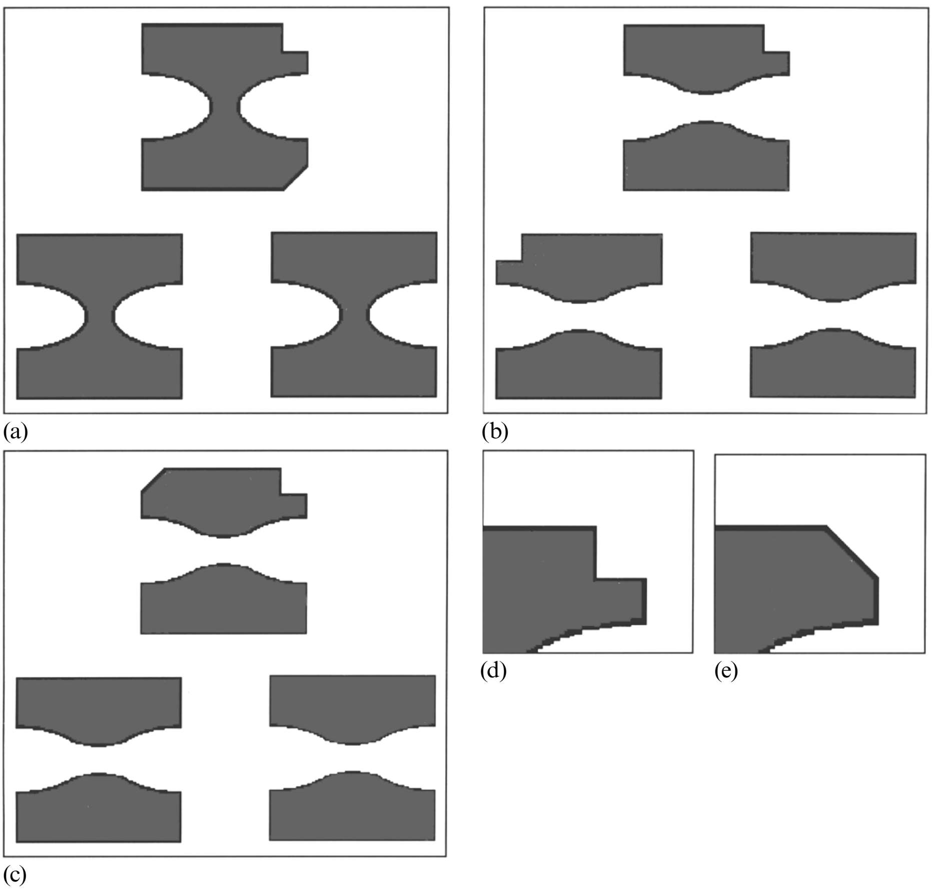

Figure 2. Stimuli employed in experiment 1. (a) Typical three-object display in which the target features are vertically separated and appear on a single large object. In this case the two features are of different types (one square notch, one diagonal). (b) Six-object display in which the two target features (square notches) appear on two distant objects. In this case the two features are of the same type (both square notches). (c) Six-object display where target features (one diagonal notch, one square notch) are horizontally separated and appear on a single small object. (d) and (e) Magnified examples of square and diagonal notch target features.

are drawn to scale—the actual stimuli measured 20 cm vertically and were viewed from a distance of 50 cm. From this dimension, all other stimulus dimensions can be calculated given the scaled figure. A feedback symbol (+ or −) immediately followed each response, appearing centrally and subtending 0.5 deg of visual angle.

Note that, as for the Davis et al study, the displays were presented only in their illustrated orientations for this study (see figures 1d – 1f and figure 2) and not as other possible versions in which each large object or pair of small objects would be rotated by 90°. This was because pilot observations indicated that, when pairs of objects in the six-object displays were constructed with one object on the left and one on the right, the gap between those objects was seen as symmetrical around its major axis, the vertical axis. This symmetry caused the white gap to be seen as a vase-like single figure, and the dark objects to become ground, which was undesirable given the importance of controlling the number of figural objects seen. As an additional precaution against the effects of figure – ground reversal, all our studies employed stimuli where the target notches appeared on the outer corners of each object. This ensured that even if a vase-like white figure were perceived in the centre of each object, the features could not be perceived to pertain to this vase, only to the dark objects themselves (see eg figure 1e).

2.1.3 *Procedure.* Target features could be square (see figure 2d for magnified example) or diagonal (see magnified example in figure 2e) and were equally likely to appear at any of the four possible 'corner' locations. Three-large-objects displays comprised half the trials, and six-small-objects displays the other half. The order in which different trial types were presented was randomised. Each observer viewed ten blocks of sixty trials, the first four blocks of which were excluded as practice.

Observers were instructed to focus their gaze at the centre of the screen, and to try to attend all of the objects presented simultaneously. When the target features were presented, observers had to press the correct button, as quickly and accurately as possible, to indicate whether the two notches were the same or different in shape. Observers pressed one key on a computer keyboard when the two notches had the same shape (ie both diagonal, or both rectangular) and another key when the notches had different shapes (ie one diagonal, one rectangular), which was equally likely. Following each response, a +/− feedback symbol was presented.

2.2 *Results and discussion*

Trials where the features were the same (both rectangular or both diagonal notches) or different (one diagonal, one square) were pooled so as to yield a minimum of thirty trials per cell for all possible comparisons. Overall means of median RTs were calculated for the three-large-objects conditions (678 ms) versus six-small-objects conditions (682 ms). These data were analysed with a two-way mixed ANOVA [number of objects (three versus six, within subjects) × display duration (224 ms versus until response, between subjects)], which yielded no significant main effect of the number of objects ($F_{1,13} = 0.36$, ns), indicating no overall advantage for three-object versus six-object displays. Moreover, this analysis showed no effects or interactions involving notch duration (both Fs < 0.1, ns). This analysis suggested that the three-object and six-object displays had been attended equally efficiently, and that saccades had not played a major role in this finding (from the absence of significant display-duration effects).

However, the absence of any difference between six-object and three-object displays could in principle have arisen because observers perceived our six-object displays to comprise only three objects, with each neighbouring pair of objects forming a single Gestalt. To ensure that this was not the case, we performed an identical analysis of individual trial types to that calculated by Davis et al (2000) for their one-object versus two-object displays. For six-object displays we compared RTs for the subset of trials in which two features had appeared on a single small object (ie were horizontally displaced, as in figure 2c) with those for the subset of trials in which two features had appeared on two separate neighbouring objects (were vertically separated appearing on separate objects but within a single 'pair' of objects, as in figure 3). If the six objects in our six-object displays had all been perceived as three sets of two separate objects, we should expect to find the same patterns of performance as Davis et al had found for their two-object displays. That is, better performance when the features were horizontally separated, appearing on a single object, relative to when the features were vertically separated, appearing on separate neighbouring objects. Such a pattern is expected for pairs of small objects (see our description of Davis et al in section 1). Next, for three-object trials, we conducted an identical comparison of horizontally versus vertically displaced features, except that they now always appeared upon a single large object.

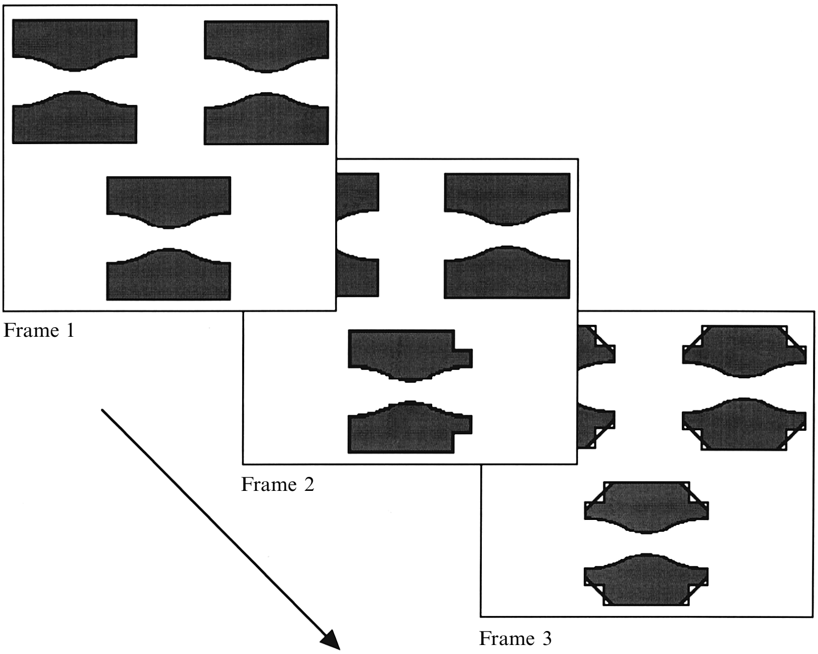

Figure 3. Schematised illustration of a typical trial's time course. Frame 1: objects are presented without target features for 2.4 s to permit observers to focus their attention effectively upon the objects. Frame 2: notches are briefly removed from objects either for 224 ms before being masked (eight observers) or until a response is made (seven observers). Note that in this example, the two target features (square notches) are vertically separated within a single pair of objects, with one feature falling on each of the objects. Frame 3: masking display that followed brief presentation (224 ms) of notches, remaining until response.

Means of median RTs for these four conditions are graphed in figure 4a (with error rates in parentheses). Filled symbols denote data points for three-object conditions, whereas open symbols denote data points for six-object conditions. The left pair of data points indicates horizontally displaced targets, and the right pair indicates vertically displaced ones. Inspection of figure 4a suggests that different patterns arose for pairs of small objects in the six-object displays than for single large objects in the three-object displays. Horizontally displaced features (same object) appear to have been detected faster than vertically displaced features (different object) for the six-small-objects displays, while no such pattern held for the three-large-objects displays.

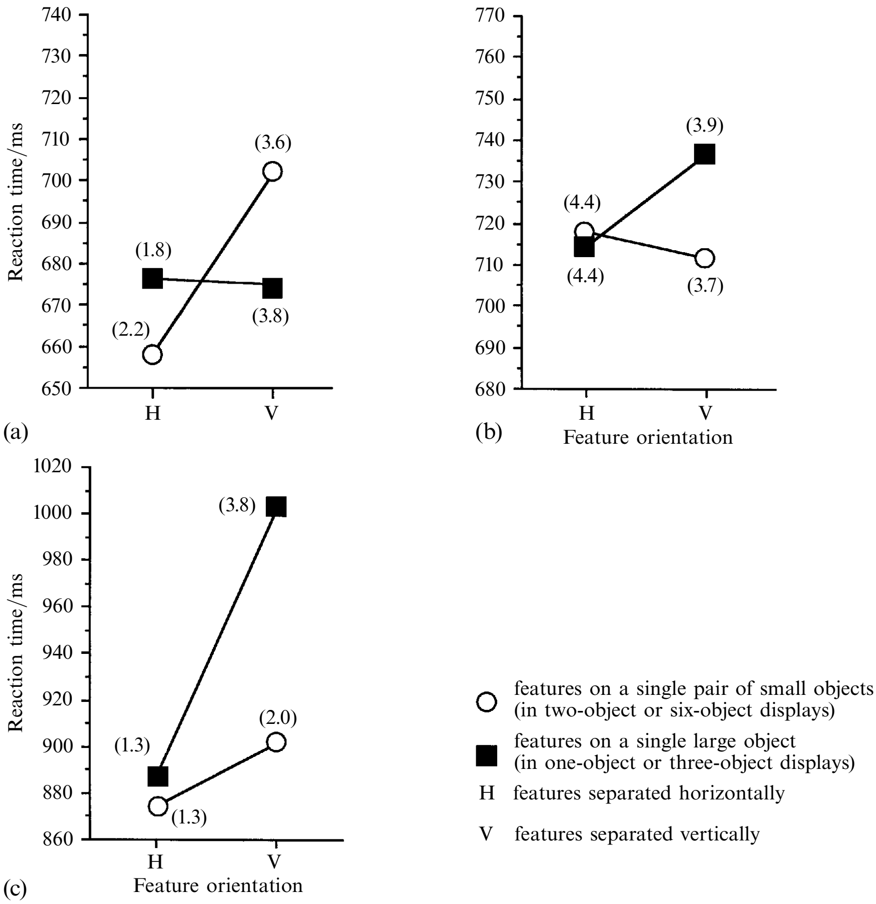

Figure 4. Graphs of RT data from current studies (percentage error rates in parentheses). (a) RT data for conditions in experiment 1 where both target features appeared on a single large object in the three-object displays or within a pair of small objects in the six-object displays. Note the faster RTs in six-object displays for horizontally displaced features belonging to the same small object than for vertically displaced features belonging to different, neighbouring objects. This pattern did not hold for horizontally versus vertically separated features belonging to a single large object in three-object displays. (b) RT data for conditions in experiment 2 for one-large-object and two-small-object displays. Note that vertically separated features within the two-object displays, which belong to separate objects, are more rapidly compared than vertically separated features belonging to the same large object. However, there is no difference between RTs for horizontally separated features in the one-object and two-object displays where target features belonged to the same perceptual object in both cases. (c) RT data for conditions in experiment 3 for features appearing on a single large object in the three-object displays and two-small-object displays. Note that vertically separated features within the two-object displays, which belong to two separate small objects, are more rapidly compared than vertically separated features within the one-large-object displays, which belong to the same large object. However, there is no difference between RTs for horizontally separated features belonging to a single large object in the three-object displays compared with those on a single small object in the six-object displays where target features belonged to the same perceptual object in both cases.

These RT data were analysed with a three-way mixed ANOVA [display type (features on a single large object in the three-object displays versus on a 'pair' of neighbouring objects in the six-object displays) × horizontal or vertical target displacement × display duration (224 ms versus until response)] that yielded no main effect of display type ($F_{1,13} = 0.26$, ns) and a main effect of horizontally versus vertically displaced targets ($F_{1,13} = 9.88$, $p < 0.01$). Crucially, however, these two factors interacted significantly

($F_{1,13} = 16.90$, $p < 0.02$). Planned comparisons revealed the source of this interaction. For features appearing within a single pair of objects in the six-small-objects displays, vertically displaced targets (which pertained to two different small objects) were detected more slowly than horizontally displaced targets (both of which pertained to the same small object); ($t_{14} = 4.7$, $p < 0.01$). This effect is typical for pairs of objects (see our description of Davis et al 2000, in section 1). However, this advantage for horizontally over vertically separated features did not arise in the three-object displays, where the opposite pattern of results held: RTs in those displays were numerically *quicker* for vertically displaced than horizontally displaced features, though this did not reach significance ($t_{14} = 0.2$, ns). Together these results, which paralleled the Davis et al findings in one large object versus two small objects, indicated that the six-object displays had been treated as comprising six separate objects (ie three pairs of separate objects), while the three-large-objects displays had been perceived to comprise only three objects. This finding validated our comparison of attention to six versus three objects, which indicated that six objects may be attended as efficiently as three. No significant terms involving display duration were yielded, indicating that saccades from one feature to another had not played a role in the current findings (all Fs < 2.3). Similar analyses of error rates also found no significant main effects or interactions that could threaten our interpretation of the RT data (all Fs < 1).

These new findings seem to provide some initial evidence against a four-object limit on visual attention: in our two-target measure, six objects were attended as efficiently as three objects. Additionally, we were able to demonstrate that the six-object displays were treated as comprising six separate objects, from the differential object-segmentation effects on performance in the two display types. This substantial effect of object segmentation also indicated that observers had not attended to the target features in isolation, but had rather been compelled to attend the objects the features belonged to.

One further possible objection might be raised against this new study. Our analyses thus far could not test whether observers had fully attended the three-object and six-object displays whilst waiting for the target features to appear. Instead of attending the objects as we had instructed, observers may have ignored them, waiting until the onset of target features 'cued' their attention to just the relevant objects in each case. However, the data seem to argue against such an interpretation.

Evidence from both Davis et al (2000) and from experiment 1 suggests that this greater amount of information to be attended in the three-object displays on such an account would have yielded significant performance deficits for the three-object displays. In experiment 1, for example, when two features appeared on one large object (horizontally separated in the three-object displays) they were *less* efficiently compared than two features on a small object (horizontally separated in our six-object displays) as tested in a two-way mixed ANOVA with feature duration as a between-subjects factor and small versus large object as a within-subjects factor ($F_{1,13} = 5.71$, $p = 0.033$). Thus, when attention needs only select a small amount of information (one small object), this makes the task easier than when attention must select a larger amount of information (a large object).

Consider how this would affect processing of six-object and three-object displays from experiment 1, if observers had only attended the object(s) on which the features appeared, rather than attending the whole display. In the three-object displays, apart from the relatively few trials (33%) in which features appeared on the same large object, they always belonged to two separate *large* objects. However, in the six-object displays, except for the few trials when features did not appear on a single small object or a single pair of neighbouring objects in the six-object displays (33%), features appeared on two separate *small* objects (see eg figure 2b). Thus, if observers had only attended the objects on which features appeared, they would for the majority of trials have attended much more information overall (two large objects) in the three-object displays than in the six-object displays (when they would attend only two small objects). If observers had attended only the objects where features appeared, attending more information in the three-object displays (two large objects) than in the six-object displays (two small objects), we should expect performance to suffer in the three-object relative to the six-object displays. Such a difference was not found, suggesting that observers had not adopted this strategy.

Our new findings provide a first indication that no four-object limit holds for visual attention as, when the overall amount of 'information' is held constant (experiment 1), six objects can be attended as efficiently as three objects. These results, and those of Davis et al (2000), suggest that visual attention is not subject to a four-object *or* a one-object limit. Indeed, it would appear from these studies that the efficiency with which we can attend a given stimulus is unaffected by the number of objects constituting that stimulus, but rather is a function of the overall amount of information comprised by the attended objects. However, the intuitive notion we have thus far employed of the 'amount of information' in a display is too vague to be manipulated experimentally. In the remaining sections of this paper we therefore describe and test a more detailed proposal regarding object-based limitations on what we can attend.

3 A feature-binding account of attention limitations

In section 1, we briefly discussed the need for vision to 'bind' together multiple visual features (fragments of shape, colour, texture, etc) to form 'objects' that our attention can select between. Although references to binding are now commonplace in the literature, there is often no explicit description given of what is meant by the term. Sometimes (eg Humphreys 1998), binding between two elements can refer to a spatial code specifying where one element lies relative to another. For our current purposes, however, binding between any two elements refers to processes that result in the two elements being selected concurrently by attention.

This notion of binding between pairs of features can help us to explain typical findings from two-target studies. When two features belong to the same object, they should be strongly bound together by early stages of vision such that they can be attended as a single unit. In contrast, binding between features from different objects may be weaker or nonexistent. If we then assume that the efficiency with which observers can attend (and thus compare) two features depends upon the strength of binding between those features, this would correctly predict that features belonging to the *same* object will often be more rapidly compared than features from separate objects.

However, it is less clear how binding can help us to understand another key finding from Davis et al (2000)—that two objects can *overall* be attended as efficiently as one object, once the amount of information is held approximately constant in the two cases. One possibility, which we consider here, is that the *amount* of binding needed to code attended objects, not the number of objects per se, is the crucial factor limiting what we attend. On such a view, as the two-object and one-object displays employed by Davis et al comprise approximately the same number of features, the same amount of feature binding overall may arise in the two cases, predicting correctly they should thus be attended equally well. The same argument can be used for the three-object and six-object displays from our experiment 1, which also presumably comprised approximately the same number of features overall and thus should also have been (and were) attended equally well.

In order to develop these proposals, we will adapt some terms and concepts from the work of Humphreys (eg Humphreys 1998; Humphreys and Riddoch 1994). Specifically, Humphreys has termed the binding that arises between any two features from the same object a 'within-object link'. In contrast, he has suggested there is no direct binding between features from separate objects. Rather, within-object links bind together features into objects and a second type of binding permits up to four of these objects to be attended at once. Humphreys referred to this second type of binding as 'between-object links', each 'link' binding together two whole objects.

Our own view differs from that of Humphreys (1998) in two primary respects. First, we suggest that there *is* direct binding between features from separate objects. That is, each between-object 'link' binds a pair of features from different objects in the same way as each within-object link binds a pair of features within the same object. Such a view could make functional sense when we are judging, for example, whether two moving objects will impact each other, or when we are guiding our fingers to grasp an object. In both cases, it is crucial to apprehend the relationships between features from *separate* objects, a process that we suggest would work most efficiently if these features are bound together and attended simultaneously.

Second, whereas Humphreys proposes that our attention can select only a fixed number of objects at a time (four), we suggest that object-based attention is limited only by the amount of binding required to code attended objects. Our alternative suggestion derives from findings in experiment 1 that seemed to provide evidence against Humphreys' view, as six objects there were attended as well as three.

A further aspect of Humphreys' (1998) account that is similar to our own view concerns the proposed neural underpinnings of within-object and between-object links. Humphreys has suggested that between-object links are coded primarily within the dorsal stream of human vision, and within-object links coded in the ventral stream. Our approach essentially adopts this view but emphasises processing at earlier stages of vision. Specifically, we suggest that magnocellular processes in areas V1/V2 (layers 4Cα and 4B/thick stripes) code between-object links, whereas parvocellular processes (layers 4Cβ, and particularly interblob regions of layers 2 and 3/pale stripes) code within-object links. Our approach can thus be summarised in terms of two claims.

(i) Individual pairs of features belonging to the same object are bound together by within-object links, driven primarily by parvocellular processes, while pairs of features belonging to different nearby objects are bound by similar between-object links, driven by magnocellular processes.

(ii) The overall number of these links, and not the number of attended objects per se, forms the primary limitation on our visual attention.

Proposal (i) can account for the different patterns of performance within the two-object versus one-object displays of Davis et al (2000) and similar patterns in single large objects of the three-object displays versus pairs of small objects in the six-object displays of experiment 1. In their two-object displays, pairs of target features appearing on the same small object (presumably bound together by within-object links) were more efficiently compared than pairs of features appearing on two separate objects (presumably bound by between-object links). According to proposal (i), this pattern reflects the fact that the objects in each display were presented for a substantial period of time (1.6 or 2.4 s) before the onset of the target features. By this time, transient visual processes coding those objects (which respond for short durations after stimulus onset) should be responding relatively weakly, while sustained visual processes (which respond steadily for longer durations after stimulus onset) should be responding relatively strongly (see eg Livingstone and Hubel 1988). Transient processes (in the magnocellular pathway) primarily drive dorsal vision and, in our account, between-object feature binding. Conversely, sustained processes (in parvocellular channels) primarily drive ventral vision and, we suggest, within-object feature binding. Accordingly, on our amended version of Humphreys' account, we would expect that, in Davis et al (2000), within-object links driven by sustained processing of the objects should be strong and between-object links driven by transient processes should be relatively weak.

This approach therefore accounts for the RT advantages found in two-small-objects displays of Davis et al, where features belonging to the same small object (grouped by strong within-object links) were compared more efficiently than features belonging to separate, neighbouring objects (bound by weaker between-object links). Presumably, the absence of this pattern for corresponding conditions in the one-large-object displays of that study was because features were bound by equally strong within-object links in both cases.

This explanation for our findings and those of Davis et al (2000), where the object – feature presentation delays were long and sustained and parvocellular vision is dominant, is not particularly intriguing in and of itself. However, this account does make a unique prediction not found elsewhere in the attention literature when transient (magnocellular) vision rather than sustained vision is dominant. Under these alternative conditions, proposal (i) would predict that the strong magnocellular activation should result in strong between-object links, whereas relatively weaker parvocellular processes should result in weaker within-object links. If we are correct, and between-object links directly bind together features from separate objects, then under such conditions, we should therefore expect features from two separate objects to be more efficiently compared than features from the same object. Such a finding would provide strong experimental support for proposal (i).

Experimental conditions under which transient vision (and we suggest between-object feature binding) is dominant can be created simply by manipulating the object – feature presentation delay in the two-object and one-object displays of Davis et al (2000) and our task from experiment 1. Now, rather than employing a lengthy object – feature presentation delay (which emphasises the effects of sustained vision), we could present objects and features at the same time (a zero object – feature presentation delay). Thus, at the time features are being coded in each trial, magnocellular coding of the object on which they appear will still be strong (having not yet waned as was the case with lengthy delays before feature presentation). In contrast, parvocellular, sustained visual mechanisms, which are less responsive to visual onsets, should be relatively weak.

We therefore conducted two further studies, replicating the study of Davis et al (2000) described above, and experiment 1 of the current paper, but with objects and features presented at the same time. On the basis of our proposal (i), we should now expect between-object links (driven by transient, magnocellular processes) to be stronger than within-object links (driven by sustained, parvocellular processes), such that comparing features from separate objects should be easier than comparing features from the same object.

4 Experiment 2: attention to one versus two objects; transient system dominant

The logic of the current study, and the Davis et al experiment, was as follows [see also the description of Davis et al (2000) in section 1]. For the two-small-objects displays, vertically displaced target features within a ‘pair’ of neighbouring objects each belonged to a different object (figure 1f), whereas horizontally displaced target features (figure 1e) pertained to a single small object (ie the upper or lower shape). Features within the one-large-object displays could also be horizontally displaced or vertically displaced (eg figure 1d), but always belonged to the same object. If the two-object displays were really perceived to comprise two objects, and the one-object displays were perceived as a single object, different patterns of results should again be found for comparisons of horizontally separated versus vertically separated features in the two display types. However, now that transient visual processing (and hence between-object links) should be dominant, rather than sustained coding of within-object links being dominant as in the Davis et al study, we should expect to find a different pattern of results within the two display types to those found in that previous study.

As we discuss below, the new comparisons between individual conditions in our one-object and two-object displays are subtly different to those of Davis et al, although they follow the same logic. This difference is to avoid unwanted effects of possible differential attention cueing to vertical versus horizontal pairs of features in our second study. Since Davis et al (and our experiment 1) presented the objects for a substantial period of time before the features were presented, we can be fairly sure that any spatial cueing from the objects' onsets would have dissipated by the time the target features were presented. However, in experiments 2 and 3, features were presented simultaneously with the objects, such that cueing/masking properties of the objects' onsets may significantly alter performance for horizontal versus vertical pairs of features.[(3)] Such effects may arise because regions between horizontal pairs of features in both display types change their level of illumination as the objects are presented, from white (background) to grey (the objects' colour). In contrast, regions between vertically separated features remain white, as now part of the object is presented at those regions. Instead of comparing horizontal pairs versus vertical pairs of features within each display type, which would differ in terms of cueing properties, we now compared performance for vertically separated features across the two display types and horizontally separated features across the two display types. These comparisons should be unaffected by potentially different cueing/masking for horizontally separated versus vertically separated features since these effects applied equally in the one-object versus two-object displays.

Our specific predictions for individual conditions were thus as follows. As the transient visual system should now be dominant, vertically separated features should now be more speedily detected in the two-object displays (since they belonged to two separate neighbouring objects and thus should be bound by between-object links) than in the one-object displays (presumably bound by within-object links). However, no such advantage should apply for horizontally separated features in the two display types (both within-object links).

Figures 1d and 1e illustrate typical one-large-object and two-small-objects displays, respectively. In every display, two 'notches' were present and the task was to determine whether these two 'notches' were the same or different in shape; each notch was either triangular or square. Whenever one notch was removed from a vertical edge, the other notch was removed from a horizontal edge, precluding any differences in display symmetry between the various conditions (see figures 1d and 1e for examples).

4.1 *Method*

4.1.1 *Observers.* Eighteen new observers from the department subject panel were recruited. Twelve were female and six were male, and their ages ranged from 21 to 30 years, with a mean of 25 years. Each was paid five pounds.

4.1.2 *Displays and procedure.* All aspects of equipment displays and procedure were as for experiment 1, except that each display now comprised only one (figure 1d) or two (figure 1e) objects identical to those employed by Davis et al (2000). These figures are drawn to scale—the actual stimuli measured 16 cm vertically, viewed at a distance of approximately 50 cm. From this dimension, all other stimulus dimensions can be calculated given the scaled figure. Additionally the object(s) and features were now presented simultaneously, either for 224 ms (eight observers), or until response (ten observers). As for experiment 1, this was to test that saccades had played no major role in our findings.

[(3)] In fact, such differential cueing of horizontal versus vertical features *did* appear to arise in experiment 3, as indicated by an overall advantage for horizontally separated features in both three-object and six-object displays there.

4.2 *Results and discussion*

Figure 4b shows interobserver means of median RTs (with error rates in parentheses) for the one-large-object and two-small-objects conditions (filled symbols versus open symbols respectively), separately for horizontally versus vertically displaced targets. Inspection of figure 4b suggests that RTs and errors are roughly equivalent overall in the one-large-object versus two-small-objects displays with any overall advantage holding in favour of the two-small-objects displays and applying solely to vertically separated pairs of features.

The RT data were analysed with a three-way mixed ANOVA [number of objects × horizontal/vertical target displacement × presentation duration (224 ms versus until response)] that yielded a marginal main effect of horizontally displaced versus vertically displaced targets ($F_{1,16} = 3.56$, $p = 0.07$). There was a main effect of one large object versus two small objects ($F_{1,16} = 8.86$, $p < 0.01$) indicating a reversal of usual costs for attending two objects rather than one—two objects were now attended more efficiently overall than one large object. Crucially, however, these two factors interacted significantly ($F_{1,16} = 7.27$, $p < 0.02$). Planned comparisons revealed the source of this interaction. Vertically displaced targets in two-small-objects displays (which belonged to two different small objects) were judged faster than vertically displaced targets in one-large-object displays (which belonged to the same object) ($t_{17} = 25.1$, $p < 0.01$). This effect points to a reliable RT advantage for features that pertained to two different objects presumably bound by between-object links over features that pertained to a single object, presumably bound by within-object links, exactly as predicted by our new account. However, the overall advantage for two-object displays appears to selectively reflect performance in vertically separated features. For horizontally separated features in the two-object versus one-object displays, no such advantage arose ($t_{17} = 3.6$, ns).

No significant terms involving feature-presentation duration were found, indicating that saccades from one target feature to the other had played no major role in our findings (all Fs < 0.7). As for experiment 1, no error trends, main effects, or interactions were found that might alter our interpretation of the RT data.

These results provide the first evidence from neurologically normal observers for more efficient coding of feature pairs from two separate objects than from a single object. Such a finding was predicted by proposal (i) for conditions that emphasise transient, magnocellular processes in vision. In order to test whether this novel finding could be replicated in displays containing multiple objects, we next conducted a second study using zero object–feature presentation delay, but this time employing the displays and methodology from experiment 1 of the current paper.

5 Experiment 3: attention to three versus six objects; transient system dominant

Experiment 3 replicated experiment 1 precisely, except that the objects in a given trial were *not* now presented for 2.4 s before the onset of target features. Rather the target features were now presented simultaneously with the objects. In experiment 2, such conditions were found to strengthen between-object links and to weaken within-object links, and we expected the same pattern of results again.

5.1 *Method*

5.1.1 *Observers.* Fifteen new observers from the department subject panel were recruited. Ten were female, five were male, and their ages ranged from 21 to 30 years, with a mean of 26 years. Each was paid five pounds.

5.1.2 *Displays and procedure.* All aspects of displays, task, and procedure were as for experiment 1, except that the target features now appeared simultaneously with the onset of the objects. For seven observers, the target displays were presented for 224 ms and for eight observers, they were presented until response. As for experiment 1, this was to verify that saccades had played no major role in our findings.

5.2 *Results and discussion*

Overall means of median RTs (and percentage errors) were calculated for the three-large-objects conditions (941 ms, 1.5% errors) versus six-small-objects conditions (913 ms, 1.2% errors). The RT data were analysed with a two-way mixed ANOVA [three versus six objects $\times$ display duration (224 ms versus until response)], which indicated a main effect of display type ($F_{1,13} = 4.97$, $p < 0.05$). As for previous studies there were no significant main effects or interactions involving display duration, indicating that saccades from one feature to another had not played a significant role in our findings (both Fs < 1.5, ns).

Next, in order to test whether between-object binding had been stronger than within-object binding in the six-object versus three-object displays, we compared RTs for several individual trial types in six-object versus three-object displays. These analyses were identical to those for one-object versus two-object displays in experiment 2, allowing any differential cueing effects for horizontally separated and vertically separated features to be partialled out (see our introduction to experiment 2). First, as the transient visual system should now be dominant, we predicted that vertically separated features should be more speedily compared in the pairs of small objects of the six-object displays (see eg figure 3) where they were presumably bound by between-object links, than within single objects of the three-object displays where they were presumably bound by within-object links (see figure 2a). However, no such advantage should apply for horizontally separated features in the two display types (both within-object links; see figure 2c for a six-object display example).

The RT data for these individual conditions were analysed with a three-way mixed ANOVA [display type (features on a single large object in the three-object displays versus on a single object or separate neighbouring objects in the six-object displays) $\times$ horizontal/vertical target displacement $\times$ presentation duration (224 ms versus until response)]. This analysis yielded a main effect of one large object versus two small objects ($F_{1,13} = 9.92$, $p < 0.01$) and a main effect of horizontally displaced versus vertically displaced targets ($F_{1,13} = 17.23$, $p < 0.001$). Crucially, however, these two factors interacted significantly ($F_{1,13} = 15.91$, $p < 0.002$). Planned comparisons revealed the source of this interaction. As we predicted, vertically separated targets in pairs of objects within the six-object displays (that pertained to separate objects) were detected faster than vertically separated targets appearing on a single object in the three-object displays (see figure 2a for an example of this latter condition; $t_{14} = 100.8$, $p < 0.01$). This effect did not arise for horizontally separated features in the two display types ($t_{14} = 12.7$, ns). No significant interactions involving display duration or percentage errors were found that could threaten our interpretation of the RT data (all Fs < 1).

The faster detection of vertically separated features in the six-object displays (where they belonged to two separate objects) versus in the three-object displays (where they belonged to the same object, see figure 2a) provides further evidence that when the transient system is dominant, between-object links become stronger than within-object links. Moreover, this advantage selectively affected vertically separated features (where the type of hypothesised linking mechanism varied between the three-object and six-object displays), not horizontally separated features that appeared within a single perceptual object in both display types.

Why have other previous studies not found performance advantages for features belonging to separate objects over features belonging to the same object? Although this is not an appropriate forum for a lengthy discussion of other two-object findings, we suggest that there are two main reasons for this absence. First, and foremost, such a finding would previously have been hard to interpret. As we discussed in section 1, it is difficult to know which types of representation constitute 'objects' for attention in any particular circumstances. Often, we can only provide evidence that a given stimulus is one 'object' rather than two (or vice versa) when it yields a 'within-object advantage' (more efficient coding of features from the same than from different objects). On such an approach, any *between-object* advantages (more efficient coding of features from different than from the same objects) arising in previous studies may have been considered uninterpretable, and attributed to the presence of other factors in the stimuli. Second, the stimuli employed in several well-known experiments have emphasised within-object over between-object binding. Whereas some have used stimuli comprising only high spatial frequencies to which the magnocellular system is relatively insensitive (eg Duncan 1984; Lavie and Driver 1996), others have used overlapping stimuli where two objects compete for figural status (eg Behrmann et al 1998; Brawn and Snowden 2000; Duncan 1984; Lavie and Driver 1996). Finally, still other studies have employed long objects – features presentation delays as we did in experiment 1, which on our account should favour within-object binding (Atchley and Kramer, in press; Brawn and Snowden 2000). Any of these characteristics may adversely and selectively affect between-object feature comparisons, hiding evidence of between-object links.

One unexpected finding from experiment 3 that was not explicitly predicted by our proposals concerned the overall advantage for six-object over three-object displays now found. This raises the possibility that under some circumstances the *strength* of the links coding a particular display, as well as their number, may determine how well the display is attended. For example, the six-object displays in experiment 4 were presumably coded by more (strong) between-object links and fewer (weak) within-object links overall than the three-object displays. Thus, the overall average link strength may have been greater for the six-object than the three-object displays, possibly causing the six-object displays to be attended more efficiently overall.

Such an explanation would, however, leave unexplained why the greater strength of within-object links in experiment 1 did not lead to an advantage for three-object displays there, as those displays would have employed more (strong) within-object links (and fewer weak between-object links) than the six-object displays. We suggest here that when one type of link is particularly weak (as arose for within-object links in experiment 3) the overall strength of links has a significant impact upon overall performance. However, when both types of link are reasonably strong (possibly as in experiment 1), this effect of overall link strength is very small and may not affect overall performance to an extent where it can be measured in our behavioural tasks. In this latter case, only an effect of the *number* of links coding a display will be evident in patterns of performance, as suggested in proposal (ii).

6 General discussion

Experiment 1 was initially designed to resolve the debate over whether one-object or four-object limits form the primary restriction on visual object-based attention. In contrast to other two-target studies, we (and Davis et al 2000) explicitly attempted to equate overall 'complexity' across conditions, varying the number of 'objects' alone. However, under these new conditions, no evidence for one-object or four-object limits was found. Indeed, in these new studies, the number of objects did not seem to affect how efficiently a given display could be attended. To account for these findings, we detailed two new proposals based on suggestions made by Humphreys and colleagues (see eg Humphreys 1998), but different in two crucial respects. First, we suggested that there is direct binding of features from separate objects as well as from the same object. Second, we proposed that the number of within-object and between-object links, not the number of objects per se, forms the primary limitation on what we can attend.

Experiments 2 and 3 yielded some initial evidence for the first of these proposals, suggesting that under some circumstances features from separate objects can be *more* efficiently compared than features from the same object. Such a finding is not predicted by other object-based approaches to understanding visual attention, even that of Humphreys (1998) on which our own account is based. Humphreys suggested that there is no direct binding between features from separate objects, and that each between-object link binds together two whole 'objects' rather than pairs of features. In such a case, two features from the same object would be directly bound by a single within-object link, whereas two features from separate objects would first have to be bound into their individual objects by within-object links and the objects would then be bound by a between-object link. On Humphreys' view, therefore, it is difficult to see how relative weakness of within-object links (and relative strength of between-object links) would result in easier comparison of features from separate objects versus from the same object (as arose in experiments 2 and 3). Such a relative weakening of within-object links should adversely affect both within-object and between-object feature comparisons as both require within-object links. However, we propose here that there *is* direct binding of features from different objects (between-object links), which arises independently of any within-object binding. On this alternative view, a selective weakening of within-object links should only affect within-object feature comparisons, leaving between-object feature comparisons unaffected and thus predicting an advantage for between-object comparisons.

The second aspect of our proposals that differed from those arising from previous studies (eg Duncan 1984; Humphreys 1998; Pylyshyn and Storm 1988) concerned our suggestion that the number of 'links', not the number of 'objects' comprised by a given stimulus, governs how efficiently we attend that stimulus. That is, provided they comprise the same number of links overall, many objects can in principle be attended as efficiently as just one object. Thus, in answer to the question in the title of our paper, asking whether our attention can select only a fixed number of objects at a time, our view suggests the answer "no". However, whereas our experiments have provided some initial evidence that the number of objects per se in an attended stimulus does not affect how efficiently we attend it, we have as yet been unable to demonstrate a direct connection between the number of links expected in a given display and the efficiency with which that display is attended. Further research will be needed to examine this issue, but for the moment we can at least ask whether our account can in principle explain performance in previous studies.

As we mentioned in section 1, many previous enumeration and tracking studies suggest that four objects can be attended with relatively small costs compared with one or two objects, whereas attending to five or six objects yields significant costs. More specifically, the costs for attending six objects compared with four is 240 ms per item (in RT studies), whereas the costs for attending four objects versus two are around 40 ms per item: a 6 : 1 ratio. Note that the increase in the number of 'features' in a display is the same from two to four objects as from five to seven objects (and in fact is proportionally greater for two to four objects). Thus, neither the increase in the number of features alone, nor the number of objects, can explain the greater performance costs for seven versus five items compared with four versus two items. Conventionally, therefore, such patterns of performance have been interpreted as evidence that three or four objects can be attended simultaneously, but that when five or more objects must be attended, multiple attention 'fixations' are required. However, we have rejected this conventional explanation here, as our studies have shown no evidence of a four-object limit in displays where the amount of information is kept constant. Rather, we suggest that although neither the number of features in a display or the number of objects can easily explain performance in such experiments, the number of 'links' can.

Consider the multiple-item displays in figure 5. For the sake of simplicity we shall assume that each item (filled circle) constitutes one 'feature' (though this is not necessary for our account to work). The lines between each pair of items in those displays represent individual links (within-object or between-object) that would be expected to arise between the items. Note that whereas two simple, small objects might be connected by a single link, three objects would tend to generate three such links (one for each possible pairing of objects in the display). For four objects, six links would be formed, for five, ten links, and so on. The number of links increases more and more rapidly as the number of attended items increases, just as RT costs do. When we calculated the number of links squared in such displays $[(n(n-1)/2)^2$ where n is the number of items], we found that this increased approximately 5.4 times as rapidly in displays of six versus four objects as in displays of four versus two objects, approximating the 6 : 1 ratio of RT increments found in enumeration and tracking studies.

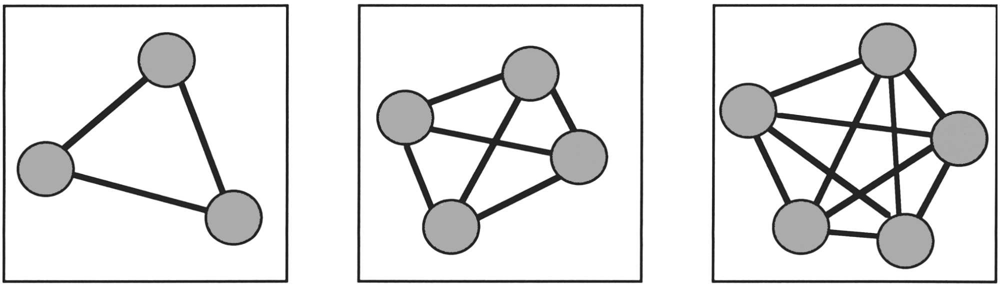

Figure 5. Illustration of hypothesised links for displays comprising 3, 4, or 5 items, respectively. The number of links (binding features from separate circles, indicated by black lines) increases disproportionately to the number of objects (3, 6, and 10 links for 3, 4, and 5 objects, respectively).

This simple analysis suggests that, by considering the number of links required to bind features of attended objects, we can account for patterns of performance in previous tracking studies. In our illustration here, we have assumed that there is one link between each pair of items. However, this is only assumed for the sake of simplicity in our illustration of the model; similar predictions could be derived if more than one link arose between each pair of items. Similarly, it does not matter in our view whether all the links binding such items are within-object, all are between-object, or whether there are some of each.

In summary, the arguments and empirical findings presented here suggest that there is no fixed number of 'objects' that we can attend at any one time. Rather, when the overall amount of information is held constant, the number of objects comprised by attended stimuli does not affect the overall efficiency with which we attend those stimuli. A second finding arising in the current studies suggests that there is direct 'binding' together of features from separate objects (possibly by magnocellular processes in V1/V2) as well as from the same object (possibly by parvocellular processes). When the likely amount (and strength) of this binding is calculated for a given set of stimuli, this can readily explain results in previous two-target, enumeration, and tracking studies.

Acknowledgements. This work was funded by the Royal Society Dorothy Hodgkin Fellowship and a joint DERA/EPSRC grant awarded to the first author. We would like to thank John Duncan and Glyn Humphreys for their comments on the manuscript.

References

Atchley P, Kramer A (in press) "Object-based attentional selection in three-dimensional space" *Visual Cognition*

Behrmann M, Zemel R, Mozer M, 1998 "Object-based attention and occlusion: evidence from normal participants and a computational model" *Journal of Experimental Psychology: Human Perception and Performance* **24** 1011 – 1036

Brawn P, Snowden R, 2000 "Attention to overlapping objects: detection and discrimination of luminance changes" *Journal of Experimental Psychology: Human Perception and Performance* **26** 342 – 358

Davis G, Driver J, Pavani F, Shepherd A, 2000 "Reappraising the apparent costs of attending to two separate visual objects" *Vision Research* **40** 1323 – 1332

Duncan J, 1984 "Selective attention and the organisation of visual information" *Journal of Experimental Psychology: General* **113** 501 – 517

Duncan J, Humphreys G, Ward R, 1997 "Competitive brain activity in visual attention" *Current Opinion in Neurobiology* **7** 255 – 261

Egly R, Driver J, Rafal R D, 1994 "Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects" *Journal of Experimental Psychology: General* **123** 161 – 177

Enns J T, Rensink R A, 1992 "Vscope software and manual: vision testing software for the Macintosh", Micropsych Software, Vancouver, Canada

Humphreys G W, 1998 "Neural representation of objects in space: a dual coding account" *Philosophical Transactions of the Royal Society of London, Series B* **353** 1341 – 1351

Humphreys G, Riddoch M J, 1994 "Attention to within-object and between-object spatial representations: multiple sites for visual selection" *Cognitive Neurology* **12** 283 – 312

Lavie N, Driver J, 1996 "On the spatial extent of attention in object-based visual selection" *Perception & Psychophysics* **58** 1238 – 1251

Livingstone M, Hubel D, 1988 "Segregation of colour, form, movement and depth: anatomy, physiology and perception" *Science* **240** 740 – 749

Pylyshyn Z, 1989 "The role of location indices in spatial perception: A sketch of the FINST spatial-index model" *Cognition* **32** 65 – 97

Scholl B J, Pylyshyn Z W, 1999 "Tracking multiple objects through occlusion: clues to objecthood" *Cognitive Psychology* **38** 259 – 290

Sears C R, Pylyshyn Z W, 2000 "Multiple object tracking and attentional processing" *Canadian Journal of Experimental Psychology* **54** 1 – 14

Trick L, Pylyshyn Z, 1993 "What enumeration studies can show us about visual attention: evidence for limited capacity preattentive processing" *Journal of Experimental Psychology: Human Perception and Performance* **19** 331 – 351

Watson S E, Kramer A F, 1999 "Object-based visual selective attention and perceptual organization" *Perception & Psychophysics* **61** 31 – 49

© 2001 a Pion publication printed in Great Britain