ELSEVIER

Cognition 80 (2001) 1–46

COGNITION

www.elsevier.com/locate/cognit

Objects and attention: the state of the art

Brian J. Scholl*

Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520-8205, USA

Received 31 January 2000; accepted 17 November 2000

Abstract

What are the *units* of attention? In addition to standard models holding that attention can select spatial regions and visual features, recent work suggests that in some cases attention can directly select discrete objects. This paper reviews the state of the art with regard to such 'object-based' attention, and explores how objects of attention relate to locations, reference frames, perceptual groups, surfaces, parts, and features. Also discussed are the dynamic aspects of objecthood, including the question of how attended objects are individuated in time, and the possibility of attending to simple dynamic motions and events. The final sections of this review generalize these issues beyond vision science, to other modalities and fields such as auditory objects of attention and the infant's 'object concept'. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Objects; Attention; State of the art

1. Introduction

In the vast literature concerning visual attention, perhaps no topic has engendered more recent work and controversy than the nature of the underlying *units* of attentional selection. Traditional models characterized attention in spatial terms, as a spotlight (or perhaps a 'zoom lens') which could move about the visual field, focusing processing resources on whatever fell within that spatial region – be it an object, a group of objects, part of one object and part of another, or even nothing at all. Recent models of attention, in contrast, suggest that in some cases the underlying units of selection are discrete visual objects, and that the limits imposed by

* Fax: +1-617-495-3764.
E-mail address: brian.scholl@yale.edu (B.J. Scholl).

PII: S0010-0277(00)00152-9

attention may then concern the number of objects which can be simultaneously attended.

This special issue of *Cognition* is concerned with the idea that attention and objecthood are intimately and importantly related. The papers in this collection review the evidence for object-based attention, discuss what attentional objects are, and discuss links both to other modalities (e.g. auditory objects of attention) and to other fields of study (e.g. developmental work on the nature of the infant's 'object concept').

1.1. Why study objects and attention?

These issues are important and timely for at least three reasons. First, the nature of the units of attention is clearly a central question for vision science: among the most crucial tasks in the study of any cognitive or perceptual process is to determine the nature of the fundamental units over which that process operates. A second reason for exploring objects and attention involves the breadth of interest in these topics: research on objects and attention has involved a convergence between many different fields of study, including experimental cognitive psychology, neuropsychology and cognitive neuroscience, philosophy of mind, developmental psychology, computer modeling, and the psychology of audition. Indeed, in a larger context, this concern for 'objecthood' can be seen as a type of 'case study' in cognitive science – an issue which is being addressed in surprisingly similar ways across traditional academic boundaries – and one of the primary goals of this special issue is to explore such connections.

A third reason for exploring these questions is that the nature of the units of attention may also prove crucial for other fields, wherein assumptions about attention frequently play a role in guiding theories of higher-order cognitive processing. As an example taken from cognitive developmental psychology, consider the following claim:

> Perceptual systems do not package the world into units. The organization of the perceived world into units may be a central task of human systems of thought… The parsing of the world into things may point to the essence of thought and to its essential distinction from perception. Perceptual systems bring knowledge of an unbroken surface layout… (Spelke, 1988b, p. 229)

The context of this historical claim is a discussion of the nature of the processes underlying various looking-time results concerning the infant's 'object concept'. The inference is that the architectural locus of these results must be 'conception', since 'perception' doesn't parse the world into units (see Scholl & Leslie, 1999 for discussion and other examples of this inference).[1] Yet, just because processing is not based on a continuous retinal layout does not necessarily mean that it has left the

[1] This type of inference continues to be influential (see Scholl & Leslie, 1999), despite the fact that many researchers in cognitive development (notably the quote's author) now take a much more nuanced view which does allow for explanations involving attention (e.g. see Spelke, Gutheil, & Van de Walle, 1995).

domain of perception. Indeed, many object-based attention results suggest that this 'packaging of the world into units' (and fairly sophisticated units at that!) may occur quite early, and even preattentively. The relation between objects and attention is thus of interest beyond vision science, and may play a role in theorizing about other cognitive processes.

1.2. A roadmap for this paper and this special issue

The goal of this paper is to review the state of the art with regard to objects and attention, and to provide a context from which the other papers in this special issue can be related to each other. This review is divided into six additional primary sections. Section 2 provides a brief review of the evidence for object-based attention, drawing on work from both experimental psychology and neuropsychology. In Section 3, these objects of attention are related to other fundamental concepts, including locations, reference frames, perceptual groups, surfaces, and parts. This section also introduces the paper by Driver and colleagues (Driver, Davis, Russell, Turatto, & Freeman, 2001), which discusses in more detail the relationship between attention and segmentation. Section 4 discusses another fundamental contrast, between objects and the individual visual features which characterize them. Section 5 discusses the dynamic aspects of objecthood, including the question of how object tokens are individuated and maintained over time. Pylyshyn's paper (Pylyshyn, 2001) focuses on this topic, and on how the earliest stages of this process serve to link up the mind and the world. This section also discusses how attention might interact directly with information which is inherently dynamic, for example simple stereotypical motions of objects. Such representations are the focus of the experiments on 'attentional sprites' reported by Cavanagh, Labianca, and Thornton (2001). Section 6 emphasizes the importance, for future work, of determining the precise properties which mediate the degree to which visual feature clusters are treated as objects. Some early work along these lines is reviewed, including the experiment reported by Scholl, Pylyshyn, and Feldman (2001a). Finally, Section 7 generalizes these issues beyond vision science, focusing on the nature of auditory objects of attention (the topic of the paper by Kubovy & Van Valkenburg, 2001), and relations to the infant's object concept (the topic of the paper by Carey & Xu, 2001).

1.3. What is attention?

Before getting to object-based attention, however, we can briefly consider a more fundamental question: what is attention, that it might be object-based? The notion of attention has been variously characterized as both obvious and intuitive, and as somehow vague and suspect. Compare:

> Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. (James, 1890, pp. 403–404)

> [P]eople talk about attention with great familiarity and confidence. They speak of it as something whose existence is a brute fact of their daily experience and therefore something about which they know a great deal, with no debt to attention researchers. (Pashler, 1998, p. 1)

> But [attention's] towering growth would appear to have been achieved at the price of calling down upon its builders the curse of Babel, 'to confound their language that they may not understand one another's speech'. For the word 'attention' quickly came to be associated … with a diversity of meanings that have the appearance of being more chaotic even than those of the term 'intelligence'. (Spearman, 1937, p. 133, quoted in Wright & Ward, 1998)

Some of the central aspects of our everyday notion of attention are reviewed by Pashler (1998): the fact that we can process some incoming stimuli more than others (*selectivity*), an apparent limit on the ability to carry out simultaneous processing (*capacity limitation*), and the fact that sustained processing of even visual stimuli seems to involve a sometimes aversive – though sometimes enjoyable – sense of exertion (*effort*). Intuitively, attention seems to be an extra processing capacity which can both intentionally and automatically select – and be effortfully sustained on – particular stimuli or activities.

The *explananda* of theories of attention are difficult to characterize precisely, and seem to comprise a family of questions related to the selectivity, effort, and capacity limitation embodied in our pretheoretical notions: why do certain events seem to automatically distract us from whatever we are doing, 'capturing' our attention? How is it that you can sometimes focus so intently on some task that you fail to perceive otherwise salient events occurring around you? Why is it that you sometimes fail to perceive clearly visible objects or events occurring right in front of you, even when you are searching for them? How is it that some activities which initially seem to require substantial effort eventually seem to become automatic and effortless? Why is it that other practiced activities do not? Why is Waldo hard to find, and how do we actually go about finding him?[2] Each of these questions has been operationalized in various experimental paradigms, many of which are reviewed below.

Because the *explananda* of attention comprise a family of 'intuitive' questions rather than a detailed operationalized problem, many people dismiss talk of 'attention' as vague or unscientific. This attitude seems misguided, however: rigor and concreteness are to be desired in scientific explanations, but cannot always be imposed on *explananda*. The questions asked above are indeed vague and hard to specify precisely, but acknowledging this does not make them go away. In this article it will be assumed that such questions are real and important, and that

[2] *Where*'s *Waldo?* is a popular series of children's activity books which embody difficult visual search tasks. Note that this is one case in which the effort involved in the allocation of attention seems to be enjoyable.

there are (possibly several different) types of selective processing – which will collectively be called ‘attention’ – that play a ubiquitous and important role in visual processing. Our topic will be the nature of the basic *units* of such selection.[3]

2. Evidence for object-based attention

In this section, some of the evidence for object-based attention is introduced. (For earlier reviews of some of this evidence, see Driver and Baylis (1998) and Kanwisher and Driver (1992).) After briefly discussing the most influential evidence for spatial selection, evidence from four experimental paradigms is reviewed (selective looking, divided attention, attentional cueing, and multi-element tracking), along with object-based phenomena in two neuropsychological syndromes (neglect and Balint syndrome).

2.1. Evidence for spatial selection

The contrast which most directly motivated the study of object-based attention was between objects and locations. Does attention always select spatial areas of the visual field, or may attention sometimes directly select discrete objects? (See Section 3.1 for a more detailed discussion of how objects and locations might be related.) The canonical evidence for spatial selection, which gave rise to the dominant ‘spotlight’ and ‘zoom lens’ models of spatial attention, comes from spatial cueing studies. Posner, Snyder, and Davidson (1980), for instance, showed that a partially valid cue to the location where a target would appear speeded the response to that target, and slowed responses when the cue was invalid and the target appeared elsewhere. A similar experiment was conducted by Downing and Pinker (1985), this time cueing one of a row of ten boxes with a partially valid cue. Detection of the targets was fastest in the cued box, and slowed monotonically as the distance between the cued box and the actual target location increased on invalid cue trials. These types of results suggested that attention was being deployed as a spatial gradient, centered on a particular location and becoming less effective as the distance from that location increased. For other spatial studies, focusing on the ‘spotlight’ and ‘zoom lens’ characterizations of attention, see the influential papers of Eriksen, Hoffman, and colleagues (Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1972, 1973; Eriksen & St. James, 1986; Eriksen & Yeh, 1985; Hoffman & Nelson, 1981) and the recent review by Cave and Bichot (1999).

2.2. Early suggestions from ‘selective looking’

Some of the earliest evidence for object-based selection came from the work of

[3] For general reviews of attention research, see Pashler (1998) and Styles (1997). For experimental phenomena which pose these broader questions of attention in a salient manner, and which highlight the difference between attention and vision more generally, see recent work on change blindness (e.g. Rensink, O’Regan, & Clark, 1997; Simons & Levin, 1997) and attentional resolution (e.g. He, Cavanagh, & Intriligator, 1997).

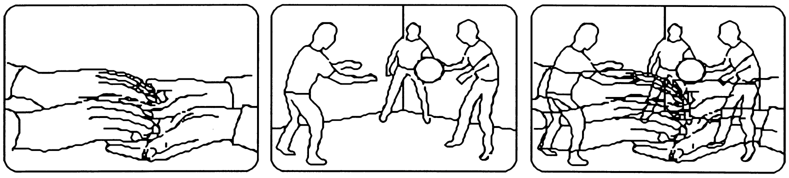

Fig. 1. Sample displays from Neisser and Becklen (1975). The two scenes – the 'hand game' and the 'ballgame' – are superimposed, and subjects are then induced to attend to only one of them, for example to count the number of times the hands clap each other. In this case, subjects fail to perceive incredible sustained events which occur in the other scene, despite the superimposition. See text for details.

Ulric Neisser on what he called 'selective looking' (Neisser, 1967, 1979; Neisser & Becklen, 1975). Subjects in these experiments simultaneously viewed two spatially superimposed movies, as in Fig. 1, and were given a 'selective looking' task which required them to attend to one of the scenes (e.g. a 'hand game', in which they had to count the number of times one set of hands hit another) and ignore the other (e.g. a ballgame, with several men passing a basketball in the background). While engaged in such tasks, these subjects failed to notice unexpected events which happened in the unattended scene (e.g. several women walking on and replacing the men in the 'ballgame' scene). By today's standards these early studies had several methodological flaws, but the essential finding – that subjects were unaware of events occurring in the unattended scene – has been replicated in more recent work (Simons & Chabris, 1999), including studies which adapted this 'selective looking' idea to computerized displays with simple shapes, wherein the details of the displays could be rigorously controlled (Most et al., in press).

This type of attentional selection seems unlikely to be spatially mediated, since the two scenes were globally superimposed: if a spatial spotlight was focused on one scene, it would also be focused on at least part of the other, and would encompass the unexpected event. As such, this early work provides evidence that attention does not simply consist of a single unitary region of spatial selection. One ironic aspect of this work, though, is that it used more naturalistic and dynamic displays than most recent studies. As discussed below, it is unclear that a movie in these experiments constitutes a single object (rather than a perceptual group, or an extended event). A ripe strategy for further research might thus involve combining the richness of Neisser's 'selective looking' stimuli with the more recent and rigorous divided attention and cueing paradigms described below.

2.3. *'Same-object advantages' in divided attention*

The type of 'overlapping' strategy used by Neisser and Becklen (1975) and others to avoid purely spatial explanations was also employed in a seminal study of divided attention by Duncan (1984) (see also Treisman, Kahneman, & Burkell, 1983). Subjects viewed brief masked displays, each containing a box with a single line drawn through it. Both the box and the line varied on two dimensions: the box could

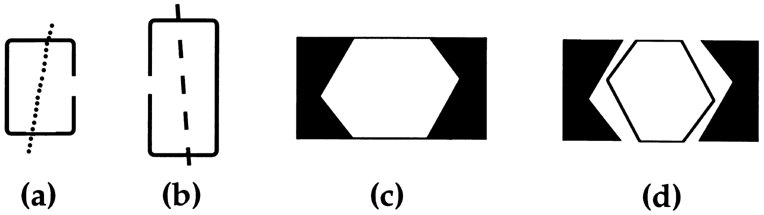

Fig. 2. Sample displays from Duncan (1984) and from Baylis and Driver (1993). (a,b) Stimuli from Duncan (1984). Each stimulus has four degrees of freedom: the line can be either dashed or dotted and can be tilted to the right or left, and the box can be either tall or short and have a right gap or left gap. Subjects are better at reporting two features from a single object compared to two features from two different objects. (c) A stimulus from Baylis and Driver (1993): subjects report the relative height of the two inner vertices while grouping the stimulus as either a single white object or two black objects (the actual colors were red and green), and are better in the single-object case. (d) An 'incongruent' control condition used to insure that subjects were grouping as they were instructed. (Adapted from Baylis and Driver (1993) and Duncan (1984).)

be tall or short, and had a small gap on either its left or the right side, and the line could be either dotted or dashed, and was oriented slightly off vertical, to either the left or the right. Fig. 2a,b present two examples of this type of stimulus. On each trial, subjects saw a brief masked box/line pair, and simply had to judge two of these properties. Some subjects were asked to make both judgments about the same object (e.g. the size of the box and the side of its gap), whereas others made a judgment about two different objects (e.g. the size of the box and the orientation of the line). As in earlier studies showing deficits for reporting two targets in a single display (Duncan, 1980), subjects were less accurate at reporting two properties from separate objects, but were able to judge two properties of a single object without any cost; this has been termed a 'same-object advantage'. Again, space-based theories cannot easily account for this result: because of the overlapped objects, the spatial extents involved in the two-object judgments were never greater than those involved in the single-object judgments.

Later studies have carried this demonstration through several iterations of proposed confounds followed by replications with the appropriate controls. Watt (1988), for instance, proposed a computational algorithm which accounted for Duncan's results in a completely data-driven manner (involving fine-grained versus course-grained spatial filters), without individuating the line and the box as different objects. A later study (Baylis & Driver, 1993) countered by using stimuli for which Watt's alternative explanation (and any such explanation based on image statistics) was inadequate: they used the same physical display for the one-object versus two-object cases, with the difference being defined by perceptual set (see Fig. 2c,d for details). Space-based theories cannot easily account for such results, since spatial location does not vary with the number of perceived objects. Note, however, that the interpretation of these divided attention tasks is still controversial because of other spatial concerns (see also Baylis, 1994; Chen, 1998; Gibson, 1994; Lavie & Driver, 1996). It has recently been argued, for instance, that the results of these divided

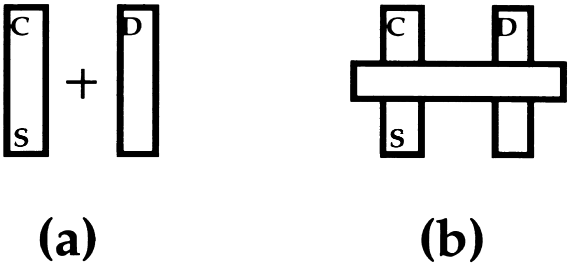

Fig. 3. Stimuli from various experiments used to demonstrate 'same-object advantages' in the automatic spread of attention: (a) Egly, Driver, and Rafal (1994); (b) Moore, Yantis, and Vaughan (1998). In each case 'C' indicates the cued location, 'S' indicates a same-object target location, and 'D' indicates a different-object target location. See text for details. Note that the Moore et al. study actually used a slightly different task. (Adapted from Egly, Driver, and Rafal (1994) and Moore et al. (1998).)

attention studies are due to the fact that automatic attentional spread has a greater area to fill with two objects than with one object, and that no same-object advantages are observed when this confound is removed (Davis, Driver, Pavani, & Shepard, 2000). The details of this interpretation still implicate object-based attention, but the mechanism responsible is seen to be automatic spread of attention, as discussed in Section 2.4. For other studies which have explored same-object advantages in divided attention tasks (some of which are discussed in later sections), see Duncan (1993a,b), Duncan and Nimmo-Smith (1996), Kramer and Watson (1996), Kramer, Weber, and Watson (1997), Valdes-Sosa, Cobo, and Pinilla (1998), Vecera and Farah (1994), and Watson and Kramer (1999).

2.4. 'Same-object advantages' in the automatic spread of attention

The studies reviewed in Section 2.3 were divided attention tasks, in which subjects attended to parts of multiple objects. Other similar studies have looked at the automatic spread of attention in response to the same type of cueing used by Posner et al. (1980) and many others to demonstrate spatial effects. Using displays such as those in Fig. 3a, Egly, Driver, and Rafal (1994) cued subjects to one end (labeled 'C') of one of two bars on each trial, using cues which were 75% valid. The subjects' task was to detect a luminance decrement at one end of a bar immediately after the cue. For the invalid cues, subjects were faster to detect targets that appeared on the uncued end of the cued bar ('S' for 'same object' in Fig. 3a), compared to the equidistant end of the uncued object ('D' for 'different object'). This is another 'same-object advantage', since the spatial distance between the cued location and the two critical locations is identical (see also He & Nakayama, 1995, described in Section 3.5).

This paradigm has also been used to demonstrate that the units of selection are at least complex enough to take occlusion into account, since the 'same-object effect' replicates with displays such as that in Fig. 3b, where the two bars are amodally

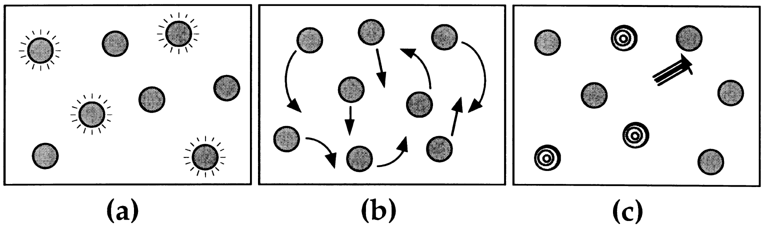

Fig. 4. A schematic depiction of a multiple-object tracking task. (a) Four items are initially flashed to indicate their status as targets. (b) All items then begin moving independently and unpredictably around the screen. (c) At the end of the motion phase, the subject must move the cursor about the screen to highlight the four targets – here the subject has just highlighted three of the targets, and is moving the mouse to the fourth. Animations of this task can be viewed or downloaded over the Internet at http://pantheon.yale.edu/~bs265/bjs-demos.html.

completed behind an occluder and so are physically separated in the display (Behrmann, Zemel, & Mozer, 1998; Moore et al., 1998). Similar results have been found for objects defined by illusory contours (Moore et al., 1998). The fact that the objects of attention in this paradigm can be defined in such ways comports with evidence from visual search paradigms that amodal completion and illusory contour formation occur preattentively (e.g. Davis & Driver, 1994; Enns & Rensink, 1998). For other studies which have explored same-object effects in the automatic spread of attention (some of which are discussed in later sections), see Atchley and Kramer (in press), Avrahami (1999), He and Nakayama (1995), Lamy and Tsal (2000), Lavie and Driver (1996), Neely, Dagenbach, Thompson, and Carr (1998), Stuart, Maruff, and Currie (1997), and Vecera (1994).

2.5. Multiple object tracking

The object-based nature of attentional selection is also apparent in dynamic situations, in which object tokens must be maintained over time. (Such dynamic objects are the focus on Section 5 of this paper.) One dynamic paradigm which has been used for this purpose is multiple object tracking (MOT), wherein subjects must attentionally track a number of independently and unpredictably moving identical items in a field of identical distractors. In the canonical MOT experiment (Pylyshyn & Storm, 1988), subjects viewed a display consisting of a field of identical white items. A certain subset of the items was then flashed several times to mark their status as targets. All of the items then began moving independently and unpredictably about the screen, constrained only so that they could not pass too near each other, and could not move off the display. At various times during this motion, one of the items was flashed, and observers pressed a key to indicate whether the flash had been at the location of a target, a non-target, or neither (see Fig. 4 for a schematic representation of this basic MOT task). Since all items are identical during the motion phase, subjects can only succeed by picking out the targets when they were initially flashed, and then using attention to track them through the motion interval.

Subjects can successfully perform this task (with over 85% accuracy) when tracking up to five targets in a field of ten identical items.[4]

Several additional results suggest that it is the items themselves which are attentionally pursued in this task as distinct objects. First, simulation results confirm that the observed performance cannot be accounted for by a single spotlight of attention which cyclically visits each item in turn, even with liberal assumptions about the speed of attentional shifts and sophisticated guessing heuristics (Pylyshyn & Storm, 1988). Second, attention has been found to speed response times to attended objects, and this advantage appears to be target-specific in MOT: in particular, it doesn't hold for non-targets, even those which are located within the convex polygon bounded by the moving targets (Intriligator, 1997; Sears & Pylyshyn, 2000). Third, as discussed below in Section 6, only certain types of visual clusters – namely those which intuitively constitute discrete objects – can be tracked in this manner (Scholl et al., 2001a). In Intriligator's terms, these results all indicate that attention is *split* between the target objects rather than being *spread* among them.[5] For other studies which have explored object-based attention with MOT, see Culham et al. (1998), Culham, Cavanagh, and Kanwisher (2001), He et al. (1997), Scholl, Pylyshyn, and Franconeri (2001b), Viswanathan and Mingolla (in press), and Yantis (1992).

2.6. Object-based neglect

'Unilateral neglect' is the name given to a collection of disorders in which patients, typically with lateralized parietal lesions, fail to perceive or respond to certain stimuli in their contralateral visual fields (for overviews, see Rafal, 1998; Robertson & Marshall, 1993). The basic phenomenon of neglect has striking practical consequences: in severe cases, neglect patients will fail to orient to people located in their neglected hemifield, will not dress the neglected side of their body, will ignore food on the neglected side of their dinner plate, etc. Historically, this class of stimuli was characterized spatially, and from an egocentric reference frame: patients were thought to neglect entire halves of their visual fields. Recent evidence,

[4] Pylyshyn himself has written of MOT as not involving attention per se, but rather an earlier preattentive tracking system (Pylyshyn, 1989, 1994, 2001). Attention, in this view, is seen as perhaps contributing to an 'error recovery' stage when a target item is 'lost', but as not being centrally involved in the tracking itself. This view is discussed at length in Pylyshyn's contribution to this special issue, but in this paper I will follow most other researchers in considering MOT as a paradigmatic case of attentional selection and attentional 'pursuit'.

[5] Yantis (1992) suggested that MOT can be enhanced by imagining the targets as being grouped into a single virtual polygon (VP), and then tracking deformations of this polygon. He demonstrated that such grouping does indeed play a role in MOT by showing that performance was facilitated simply by informing subjects of this strategy, or by constraining the items' trajectories such that the VP could never collapse upon itself. While this strategy (or, indeed, any grouping strategy, for example pairing items into virtual tumbling line segments) can indeed enhance performance, it is not necessary for successful tracking, and the enhancement seems likely to be due to an improved error recovery process when one item is lost: when items are being perceptually tracked as virtual groups, one can make an educated guess as to where a lost item 'should' be, given the overall contour of the virtual shape (Sears & Pylyshyn, 2000). In addition, Scholl and Pylyshyn (1999) have shown that dynamic information which is local to each item (or 'vertex' in the VP strategy) does greatly impact on tracking performance.

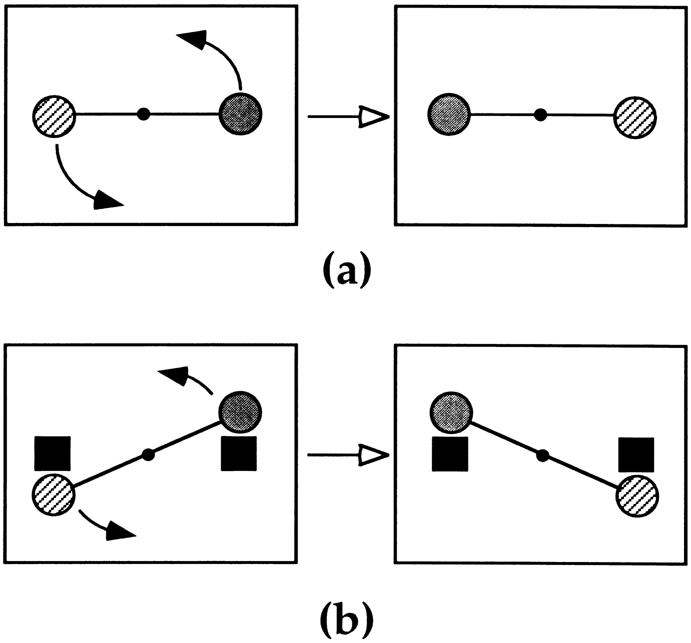

Fig. 5. Sample displays shown to neglect patients by Behrmann and Tipper (1994, 1999). (a) Subjects are shown a dumbbell which then rotates 180° in place. Left-neglect patients initially neglect the disc on the left side. After the rotation, neglect of the disk which is now on the left is attenuated, while neglect is now evident for the disk which ended up on the right. (b) The same event occurs, but two stationary boxes are added. After the rotation, left-neglect patients show some neglect for the left box but the right disc. (Adapted from Behrmann and Tipper (1994, 1999).)

however, has suggested that in some situations neglect may also be object-based, such that patients neglect entire halves of objects with salient axes regardless of the visual field in which they are presented (e.g. Caramazza & Hillis, 1990; Driver, Baylis, Goodrich, & Rafal, 1994; Driver & Halligan, 1991; Humphreys & Riddoch, 1994; Subbiah & Caramazza, 2000; Ward, Goodrich, & Driver, 1994; for many other studies, see Rafal, 1998). Here I will discuss just a single set of 'object-based' studies.

Behrmann and Tipper (1994) used a task in which left-neglect patients were required to detect targets in various 'dumbbells' consisting of two discs connected by a line. As expected, these patients were slower to detect targets presented on the left side of the dumbbell. When the whole dumbbell visibly rotated through 180°, however, these same patients then showed less neglect for the disc which ended up on the left, and were slower to detect targets presented on the disc which ended up on the right after the rotation (see Fig. 5a). Crucially, Tipper and Behrmann (1996) showed that this only held for the connected dumbbells, which were apparently treated as single objects: when the line connecting the two discs was removed, subjects were always slower to respond to targets on the left side of the display,

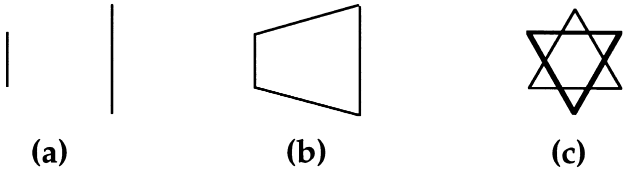

Fig. 6. Stimuli used to demonstrate object-based effects in simultanagnosic patients. Patients viewing (a) cannot determine whether the lines are of equal lengths, but they can tell that the shape in (b) is a trapezoid rather than a rectangle (Holmes & Horax, 1919). Patients of Luria (1959) could see only a single triangle from (c) at once, when the two triangles were colored differently.

regardless of their motion to or from any other location. This suggests that in some contexts these patients neglect not halves of egocentric space per se, but rather halves of specific objects with salient axes, and that the object-based reference frame which defines ‘left’ and ‘right’ persists through such rotation, such that the ‘left’ (i.e. the neglected) side of a just-rotated object can be located on the right side of the display after the rotation.

Furthermore, this type of object-based neglect can occur simultaneously with scene-based neglect (Behrmann & Tipper, 1999). When stationary squares were added to the dumbbell display, as in Fig. 5b, patients simultaneously showed some neglect for the stationary square on the left side of the display, but the dumbbell disc on the right side (to which it had rotated after being initially located on the left). This fascinating finding suggests that neglect can not only operate in multiple reference frames (including object-based ones), but can also do so simultaneously. More generally, egocentric neglect and object-based neglect may interact in other ways: for instance, the primary axis of an off-vertical object may serve to *define* egocentric left and right for an observer, such that neglect might still be considered as a primarily egocentric disorder, but with object-based contributions to the egocentric axis (e.g. Driver, 1998; Driver & Halligan, 1991).

2.7. Balint syndrome

Additional evidence for the object-based nature of visual attention comes from the study of Balint syndrome, in which (typically bilateral) parietal patients exhibit surprising object-based deficits (see Rafal, 1997, for a review). Balint syndrome, even more than neglect, is best characterized as a true syndrome, incorporating many different types of deficits which may not all share a common cause; these include near-complete spatial disorientation (including the inability to indicate an object by pointing or even by verbal description), abnormal eye movements, optic ataxia (a disorder of visually-guided reaching), and impaired depth perception. The most relevant – and startling – component of Balint syndrome, however, is termed *simultanagnosia*: the inability to perceive more than one object at a time, despite otherwise preserved visual processing, including normal acuity, stereopsis, motion detection, and even object recognition. (Such pure cases of simultanagnosia are very rare, however, and many simultanagnosic patients also have various forms of agnosia, alexia, prosopagnosia, and related deficits.) Such patients fail even the simplest of tasks which require them to compute a relation between two separate objects (Coslett & Saffran, 1991; Holmes & Horax, 1919; Humphreys & Riddoch, 1993; Luria, 1959; Rafal, 1997).

The object-based nature of this disorder emerged very early (and, indeed, seems intrinsic to the definition of simultanagnosia). Holmes and Horax (1919), for instance, noted that although Balint patients were unable to determine if two parallel lines were of equal lengths (as in Fig. 6a), they could tell whether a simple shape was a rectangle or a trapezoid (as in Fig. 6b) when the two lines were simply connected by other lines at each end to form a single shape. Similarly, although simultanagnosic patients are typically unable to see two separate discs simultaneously, they are perfectly able to see a single dumbbell (Humphreys & Riddoch, 1993; Luria, 1959). It was even noted in early work by Luria (1959) that this object-based percept seemed untied to particular locations: if the two overlapping triangles composing a 'Star of David' were colored separately, as in Fig. 6c, patients often perceived only one of them! Further aspects of Balint syndrome, involving the perception of locations and visual features, will be discussed in later sections.

2.8. Other evidence for object-based attention

Some additional evidence for object-based attention is discussed later in this article, in the context of other topics (e.g. the perception of groups, surfaces, features, and events). The goal of this review is to highlight major themes in the study of objects and attention, though, and not to exhaustively discuss the empirical evidence. As such, the rest of this article focuses on various theoretical issues and connections to other fields, and many additional studies supporting the existence of object-based selection are not discussed. These include studies of negative priming (Tipper, Brehaut, & Driver, 1990), inhibition of return (Tipper, Driver, & Weaver, 1991; Tipper, Jordan, & Weaver, 1999), symmetry judgments (Baylis & Driver, 1995; Driver & Baylis, 1996), repetition blindness (Chun, 1997; Kanwisher, 1987, 1991), attentional capture (Yantis and Hillstrom, 1994), visual illusions (Cooper & Humphreys, 1999), response competition (Kramer & Jacobson, 1991), intra-object attentional effects (Hochberg & Peterson, 1987; Peterson & Gibson, 1991), visual search (Maljkovic & Nakayama, 1996; Mounts & Melara, 1999), and visual marking (Watson & Humphreys, 1998). Other neuropsychological studies which are not discussed here include studies of early object recognition effects on scene segmentation (Peterson, Gerhardstein, Mennemeier, & Rapcsak, 1998) and suggestions of hemispheric specialization for object-based processing (Egly, Rafal, Driver, & Starrveld, 1994; Reuter-Lorenz, Drain, & Hardy-Morais, 1996).

3. Objects in context: locations, reference frames, groups, surfaces, and parts

Having now presented some of the evidence that discrete objects can in some cases serve as units of attention, it is worth stepping back from this evidence, and considering more carefully how such attended objects relate to other units and processes, including spatial locations, reference frames, perceptual groups, scene segmentation, and visual surfaces.

3.1. Objects and locations

As discussed above, the contrast which has done most to fuel research on object-based attention is between objects and locations. One general way to characterize this issue is in terms of the degree of preattentive processing in the visual system (Driver & Baylis, 1998): is an initial 'packaging of the world into units' computed before – or as a result of – attention? Viewing the question this way is much in the spirit of the classic distinction which has motivated attention research, between 'early selection' and 'late selection' theories (see Johnston & Dark, 1986; Pashler, 1998). That question typically focused on whether stimuli were processed to the level of meaning before or after the limits imposed by attention. In this context we are asking a similar question, about whether various feature clusters are parsed as independent *individuals* before an attentional bottleneck, or if the foci of attention are simply spatial in nature.[6]

It seems clear, though, that these two notions – objects and locations – should not be treated as mutually exclusive. Attention may well be object-based in some contexts, location-based in others, or even both at the same time. The 'units' of attention could vary depending on the experimental paradigm, the nature of the stimuli, or even the intentions of the observer. Perhaps attention will prove location-based *within* complex extended perceptual objects (Neely et al., 1998), or will prove object-based only under relatively distributed global spatial attention (Lavie & Driver, 1996; though see Lamy, 2000). The distinction between objects and locations may also blur in other ways, for example if the shape of a spatial spotlight is allowed to deform around an object (cf. LaBerge & Brown, 1989). It may even be, as suggested by Rafal (1997) (see also Driver, 1998; Laeng, Kosslyn, Caviness, & Bates, 1999), that object-based disorders such as simultanagnosia have their origin in disruptions of perceived space:

> A real object is perceptually distinguished from others based on its unique location; it must be in a different place from any other object. Even if it is superimposed in the retinal image, occlusion cues normally assign each of the two objects to different distances from the observer and will engender an experience of depth. Because patients lack conscious access to a visual representation of topographic space, there is only one 'there' out there – and hence there can be only one object. (Rafal, 1997, p. 350)

[6] Note that such references to preattentive processing might be a matter of degree, such that a 'preattentive' process is really best characterized as one which requires relatively little attention. It remains unclear whether there are any truly preattentive processes in the strongest sense of the term (see Nakayama & Joseph, 1998).

3.2. Object-based processing and object-based reference frames

The discussion so far has focused on the units of attentional selection. A related foundational question concerns the nature of the underlying *reference frame* into which visual features are encoded. (A reference frame here just refers to the specification of a set of axes with an origin; relational terms such as 'to the left', and 'towards the front' are then defined relative to these axes.) In an *environment-based* reference frame, visual features are encoded into some absolute coordinate system. In *viewer-based* reference frames, features are encoded relative to egocentric properties such as gaze direction or body orientation. In *object-based* reference frames, features are coded relative to axes defined by individual objects. Though it is often assumed that object-based attentional effects also implicate object-based reference frames (e.g. Behrmann & Tipper, 1999), this is not necessary (Mozer, 1999): there are many ways in which attention could spread throughout an object and not the surrounding context, even though all of the features on that object were still represented in environment- or viewer-based coordinates. One obvious way would be for the processes implementing the spread of attention to be constrained by principles of grouping, such as connectedness.[7]

The distinction between object-based processing and object-based reference frames has also been stressed (in different terms) in the context of visual attention (Vecera & Farah, 1994). Here it was noted that there are two fundamentally different ways that attention could fail to be entirely location-based. First, attention could select *groups of locations*, bound together by object formation principles (such as connectedness) but still represented in terms of their spatial coordinates; this was called a '*grouped array*' theory. On the other hand, objects could be attended without any regard for spatial position; this *spatially invariant* account is related to the idea of an object-based reference frame. (Note that while Vecera and Farah (1994) refer to only this latter type of model as 'object-based', I am considering both to be object-based accounts, since in neither case is selection based entirely on spatial location.) Which of these object-based accounts is correct appears to depend on the specific paradigm. The 'same-object advantages' in divided attention (discussed in Section 2.3) appear to reflect spatially invariant processing, since when the spatial distance between the two previously superimposed stimuli is varied, the magnitude of the object-based effect is unchanged (Vecera & Farah, 1994). In contrast, the 'same-object advantages' in spatial cueing (Section 2.4) appear to reflect the processing of grouped arrays, since manipulations of the spatial distance between the probe

[7] Mozer (1999) demonstrates this point by modeling object-based effects from the study of neglect (Behrmann & Tipper, 1999; Tipper & Behrmann, 1996, which are described in Section 2.6) in a connectionist framework which does not employ object-based reference frames. He suggests that this result supports the importance of such network models for the understanding of object-based processing, but on inspection it seems clear that all of the relevant work is done by the fact that the network operates according to the rule that, "Locations adjacent to activated locations should also be activated" (p. 458). This simply implements the connectedness constraint mentioned above, which of course could also be implemented in many other ways which did not share the details of these connectionist models.

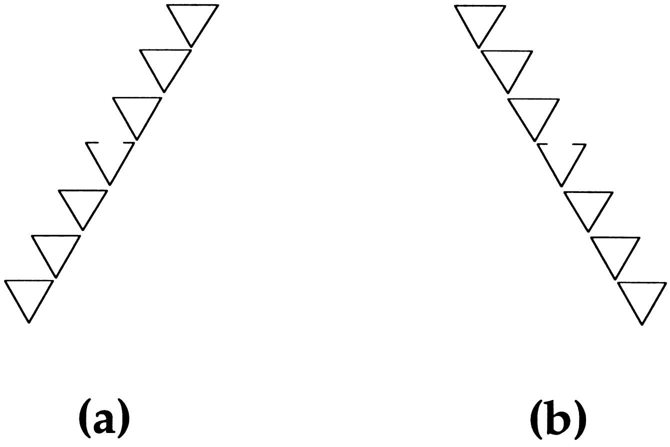

Fig. 7. Stimuli shown to neglect patients by Driver et al. (1994) to demonstrate the role of perceptual grouping in determining the axis on which the neglect is based. See text for details.

locations do attenuate the object-based effect (Vecera, 1994; though see Kramer et al., 1997; Lavie & Driver, 1996, for critiques of these conclusions).

3.3. Attention and perceptual groups

Several older research traditions have emphasized that scenes are organized into perceptual groups defined by the traditional Gestalt principles of continuity, proximity, common fate, etc. (for an excellent summary see Chapter 6 of Palmer, 1999). How does the notion of object-based attention differ from these earlier ideas? Though work on perceptual grouping has typically been conducted without reference to 'attention', many of these demonstrations could easily be reinterpreted as involving attention. For example, Driver and Baylis (1998) note that "the subjective feeling that a column or row of dots belongs together … may arise because when trying to attend to a single dot, our attention tends to spread instead across the entire group in which it falls" (pp. 301–302). Perhaps, in other words, the Gestalt psychologists were studying attention all along (see also Driver et al., 2001)!

This intriguing possibility deserves to be pursued in future work. In particular, it would be of interest to determine if the evidence for object-based selection described in Section 2 would replicate when Gestalt groups are used as stimuli instead of single objects. Some evidence suggests that it will. The 'same-object advantages' using the cueing paradigm of Egly, Driver, and Rafal (1994) (see Section 2.4), for instance, have been replicated when using two groups of circles arranged into parallel rows (Rafal, in press, cited in Egly, Driver, & Rafal, 1994; see also Driver & Baylis, 1998, pp. 303–304) – a 'same-group' advantage. In addition, some neuropsychological studies suggest a more direct role for grouping (e.g. Boutsen & Humphreys, 2000; Driver et al., 1994; Ward et al., 1994). In one study, neglect

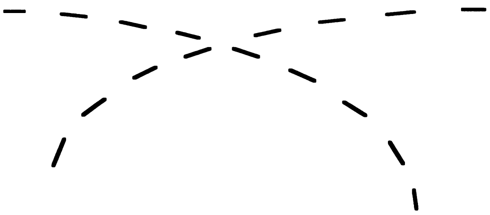

Fig. 8. A stimulus whose natural grouping is of two crossing lines, but which can also be perceived in other ways, for example as two birds' beaks touching each other.

patients reported whether a small triangle in a briefly flashed display had a gap in its contour, where this triangle was surrounded by other triangles such that it was perceptually grouped into a right-leaning or a left-leaning global figure, as in Fig. 7 (Driver et al., 1994). When the critical triangle was grouped into the left-leaning global figure, as in Fig. 7b, the gap was on the right side of this overall group; when it was perceptually grouped into the right-leaning figure, as in Fig. 7a, the gap was on the left of the overall group. This manipulation greatly affected whether the patients perceived the gap, even though the critical triangle was always drawn identically.[8]

Such evidence suggests that 'object-based' attention and 'group-based' attention may reflect the operation of the same underlying attentional circuits – a conclusion which echoes William James' comment that "however numerous the things [to which one attends], they can only be known in a single pulse of consciousness for which they form one complex 'object'" (James, 1890, p. 405). This is not a foregone conclusion, however. It may be, for example, that attention is more easily moved effortfully within *any* perceptual group that can be intentionally perceived, compared to movement between groups, but that attention will *automatically* spread only within a subset of such groups, comprising those that reflect the most 'intuitive' percepts. The line segments in Fig. 8, for example, are most naturally grouped into two crossing lines, though it is possible to perceive them in other ways, for example as two birds' beaks facing each other. Here attention might automatically spread only along the two crossing lines, despite the fact that the line segments can be grouped in several additional ways. Another way to put this is that attention may automatically spread (e.g. by 'exogenous' cues) only within groups defined primarily by 'bottom-up' factors, but that 'top-down' factors may additionally form groups which can be independently attended by intentional, endogenously-cued processes. For a more complete discussion of the relation between perceptual grouping and objecthood, see Feldman (1999).

[8] Related evidence – and, indeed, some of the earliest evidence for stimulus-based neglect – comes from studies which used words (i.e. groups of letters) as units (e.g. Caramazza & Hillis, 1990; Subbiah & Caramazza, 2000). In these cases, words seem to be treated as special cases of objects or groups (see also Kahneman & Henik, 1981).

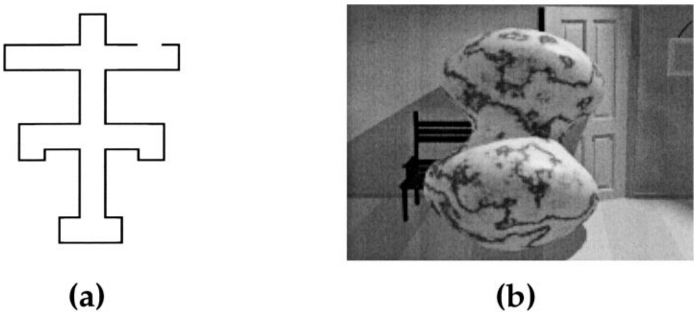

Fig. 9. Stimuli used to explore part-based attention. (a) A stimulus with intuitive 'crossbar' parts, used by Vecera et al. (2000). (b) A stimulus used by Singh and Scholl (2000) with parts defined by negative minima of curvature, the magnitude of which can be varied continuously. See text for details.

3.4. Attending to parts

Just as multiple objects can be perceptually grouped together, so can individual visual objects be composed of multiple parts (e.g. Hoffman & Richards, 1984; Palmer, 1977). The part structure of complex objects has played a major role in theorizing about the recognition of specific objects, where several researchers have proposed that specialized processes for the recognition of specific volumetric parts help to 'jumpstart' the recognition process (e.g. Biederman, 1987; Marr, 1982). In the study of attention, recent research has demonstrated 'same-part advantages' (cf. Sections 2.3 and 2.4) for complex objects composed of hierarchical part arrangements, as in Fig. 9. In one study, for example, Duncan's divided attention paradigm (Section 2.3) was tested with stimuli consisting of 'poles' with 'crossbars' (see Fig. 9a), and same-part effects on accuracy were observed concurrently with same-object effects in displays with multiple figures (Vecera, Behrmann, & McGoldrick, 2000; see also Vecera, Behrmann, & Filapek, in press).[9]

Similarly, a 'same-part advantage' was observed in a spatial cueing study (Section 2.4) with stimuli such as the one depicted in Fig. 9b (Singh & Scholl, 2000). The parts in this study were defined by minima of curvature on a 3D surface, which has been found to accurately predict where observers judge part boundaries to exist (Hoffman & Richards, 1984). This study has two advantages over the divided attention study. First, due to the nature of the experimental paradigm (based on Egly,

[9] The cueing method used in this study appears to introduce a confound, however. When subjects reported two features from a single part (e.g. whether the upper crossbar in Fig. 9a was short or long, and the side of its gap; or, alternately, whether the bottom crossbar was short or long, and the direction of its 'prongs'), they always reported all of that part's features, with the relevant part indicated by the location of a cue. In contrast, on different-part trials, subjects had to use the color of the cue to determine which feature of a part to report. This raises the possibility that the observed 'same-part advantage' simply reflects the difficulty of remembering or working through the mapping between cue color and feature-to-report in the different-part trials. In single-part trials, no such memory is required, since all features of the part are reported.

Driver, & Rafal, 1994; Moore et al., 1998), it is not subject to the confound discussed in Footnote 9. Second, defining the parts in this way allows for continuous variation in the magnitude of the curvature defining the parts, which has been found to correlate with part salience (Hoffman & Singh, 1997). Since the naturalism of the objects used here provides larger than normal cueing effects (see Atchley & Kramer, in press), it is thus possible to demonstrate that the magnitude of the 'same-object effect' varies with the magnitude of the curvature and the length of the part cuts. Both of these studies suggest that it may be worthwhile in future work to bring the literatures on attention and perceptual part structure into closer contact (Singh & Scholl, 2000; Vecera et al., 2000).

3.5. Attending to surfaces

The previous sections considered both multi-object units such as groups, and intra-object units such as parts. Visual *surfaces* constitute another level of representation which can encompass both of these categories: complex objects can consist of multiple surfaces, while multiple objects can be arrayed along a single surface. One research tradition which has been developed largely independently of the work discussed above has focused on the role of visual surfaces in 'mid-level vision' (Nakayama, He, & Shimojo, 1995). Nakayama and his colleagues argue that a surface-based level of representation is a critical link between low-level vision and high-level perception, and they have shown that several visual phenomena are based not on the retinal makeup of the visual field, but rather on the perceived interpretation of the visual field in terms of surfaces.

For example, He and Nakayama (1995) explored how attention can spread along surfaces in non-fronto-parallel depth planes. In one experiment, observers had to search for an odd-colored target in the middle depth plane of a stereoscopically presented display, ignoring the items in two other arrays at depths above and below the critical plane (see Fig. 10). When the items to be searched for were tilted so that

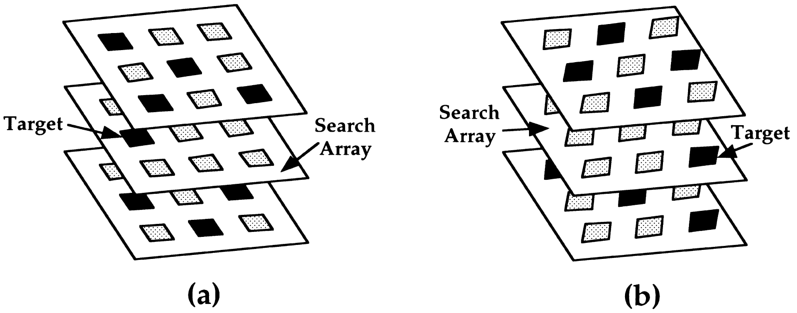

Fig. 10. Stimuli used by He and Nakayama (1995) to demonstrate that attention can efficiently select surfaces even when they span a range of depths. Subjects must detect an odd-colored item in the middle depth plane in each case. When the items are arrayed along a perceptual surface corresponding to this depth plane (a), the search is efficient. When the items do not lie along such a surface (b), attention can no longer select that depth plane, and the search is impaired. (Adapted from He and Nakayama (1995).)

they appeared to lie along a surface at this middle depth plane (Fig. 10a), subjects were able to efficiently confine their search to those items, speeding search. However, when the items were tilted so that they were not seen to lie along such a surface (Fig. 10b), subjects could no longer confine their search to the middle depth plane, and response times increased. This indicates that attention can efficiently select individual surfaces, even when they span an extreme range of depths. In another experiment, He and Nakayama used a cueing study (very similar to that of Egly, Driver, & Rafal, 1994) to demonstrate that in some cases attention *must* spread along surfaces. Subjects had to detect a target in one of two rows of items, and a cue indicated the row which was 80% likely to contain the target. When the disparity of the two rows was increased, observers were able to more efficiently select the cued row, but only if the items did not lie along a common perceptual surface; when the items did lie along a common surface, increased disparity between the rows was unable to facilitate selection of the cued row. He and Nakayama conclude that "The visual system … can direct selective attention efficiently to any well-formed perceptually distinguishable surface" (p. 11155), and while they do not explain 'well-formedness', they hint that local co-planarity and collinearity of surface edges may play important roles. These roles are considered again in Section 6 below.

As with perceptual groups, it is possible that attending to objects, surfaces, and parts all reflect the operation of the same underlying attentional circuits. Future work along these lines must investigate the extent to which phenomena of object-based attention will replicate with such units, and must also take care to pursue rigorous ways of distinguishing surfaces, objects, and parts, rather than relying on intuitive conceptions of such units (see Feldman, 1999).

3.6. Attention, segmentation, and proto-objects

As the previous three sections have emphasized, there may be a hierarchy of units of attention, ranging from intra-object surfaces and parts to multi-object surfaces and perceptual groups. It remains an open question whether attention to each of these levels reflects the operation of the same or distinct attentional circuits. Another complication is that each of these units may be computed multiple times within the course of visual processing. In general, segmentation processes – that is, processes that bundle parts of the visual field together as units – probably exist at all levels of visual processing. Some of these processes are early, using 'quick and dirty' heuristics to identify likely units for further processing. This results in a visual field which has been segmented into 'proto-objects', which are thought to be volatile in the sense that they are constantly regenerated rather than being stored in visual working memory (VWM) (Rensink, 2000a,b).

In this scheme, it is these 'proto-objects' which serve as the potential units of attention. Then, once a proto-object is actually attended, additional object-based processes come into play. In Rensink's 'coherence' theory (Rensink, 2000a,b), attention to a proto-object gives rise to a more detailed representation of that unit, and one that persists in VWM. It seems likely, however, that this attentional processing could in some cases override the earlier parsing characterized by the proto-objects. For instance, the additional processing which results from attention to a proto-object may result in a higher-level representation of that portion of the visual field as a pair of intertwined objects, or as only a part of a more global object or group of objects. In general, since such processes can occur at multiple levels, 'segmentation' cannot be considered as synonymous with object-based attention. The units of some segmentation processes may serve as the focus of attention, while the units of other segmentation processes may be in part the *result* of (proto-)object-based attention.

The relation between attention and segmentation is treated at length in the paper by Driver and colleagues in this special issue (Driver et al., 2001). They address the issues discussed above, and focus on the question of whether – and how much – image segmentation occurs with attention, without attention, and with attention otherwise occupied in a competing task. In experiments studying a wide array of phenomena – transparency, change blindness and inattentional blindness, modal and amodal completion, and low-level flanker tasks – Driver and colleagues stress how unattended (and even un*seen*) portions of the visual field can still enter into segmentation processes, while at the same time attention can influence even very early types of segmentation. Throughout this work, Driver and colleagues discuss the conscious phenomenology of scene segmentation, and the limited degree to which it represents the complexity of visual processing.

4. Objects and features

In the previous sections we considered hierarchical objects, and the possibility of attending to individual parts and surfaces. Objects are also seen as comprising individual features, however, such as color, luminance, shape, and orientation. In this section object-based selection is contrasted with feature-based models, in which attention can select individual visual features, and in which the limits imposed by attention may concern the number of such features which can be simultaneously encoded into VWM.

One recent experiment which highlights this contrast used a change detection paradigm to demonstrate that the units of VWM are in some cases discrete objects, apparently regardless of the number of visual features which make up those objects (Luck & Vogel, 1997; see also Irwin & Andrews, 1996). On each trial, subjects saw a display such as that in Fig. 11a for 100 ms, followed by a brief blank delay and then Fig. 11b for 2000 ms, and simply had to determine whether there had been a change. Using simple features such as colored boxes and oriented lines, VWM was found to have a capacity of four features, as evidenced by change detection accuracy. Surprisingly, this same limit of four discrete objects held whether the items were colored boxes (Fig. 11a,b) or oriented lines of different colors (with two features per object), or even colored oriented bars which came in two possible sizes and which either did or did not have a gap (see Fig. 11c,d), in which case all 16 features from the four objects could be retained as accurately as only four features from four objects. It thus

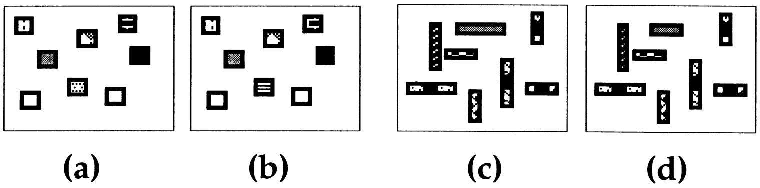

Fig. 11. Sample change detection displays from Luck and Vogel (1997), with texture standing in for color. Displays (a) and (b) contain colored boxes. Displays (c) and (d) contain bars of two different sizes and orientations and several colors, and can either have a gap or no gap. Subjects are shown display (a) for 100 ms, followed by display (b) after a 900 ms delay, and must indicate whether a change occurred (here it did: one item changed 'color'). A similar method is used for displays (c) and (d), where one of the items has changed size. In each case, subjects are accurate with displays containing up to four items in total, regardless of the number of features comprising those items. See text for details. (Adapted from Luck and Vogel (1997).)

appears that object-based processing can trump feature-based processing as well as purely spatial processing, at least in some circumstances and for some 'objects'.[10]

4.1. The link between object-based attention and feature encoding

The studies of VWM described above support a view which is often taken as a hallmark of object-based selection: that attending to an object automatically entails encoding all of its features into VWM. This thesis was proposed in early theorizing about object-based attention (e.g. Kahneman & Henik, 1981), and remains pervasive today (Duncan, 1993a,b; Duncan & Nimmo-Smith, 1996; O'Craven, Downing, & Kanwisher, 1999). Kahneman and Henik (1981), for instance, suggested that, "Attention can be focused narrowly on a single unit, or else it can be shared among several objects. To the degree that an object is attended, however, all its aspects and distinctive elements receive attention. An irrelevant element of an attended object will therefore attract – and waste – its share of attention." (p. 183). More recently, O'Craven et al. (1999) have suggested that "the central claim of object-based theories" is that "task-irrelevant attributes of an attended object will be selected along with the task-relevant attribute, even when these attributes are independent" (p. 585).

Converging evidence for this view comes from a recent neuroimaging study (O'Craven et al., 1999). Previous neuroimaging studies have identified a part of the fusiform gyrus which responds selectively to faces (the 'fusiform face area' or

[10] More recent work with this paradigm supports the existence of an object-based component to VWM, but with two important limitations. First, whereas Luck and Vogel (1997) obtained an object-based result even for objects defined by a conjunction of two identical dimensions (e.g. a colored border around a colored box), later studies have failed to replicate this effect when the second display could not contain any entirely new colors (Wheeler & Treisman, 1999; Xu, 2001b). Second, the types of 'objects' that enjoy this effect may be highly constrained. An attenuated object-based effect with color and orientation is found for colored oriented bars, for example, but not for colored 'beach balls' with colored oriented stripes running through them (Xu, 2001a).

FFA; Kanwisher, McDermott, & Chun, 1997) and also a part of parahippocampal cortex which responds selectively to the shape of the local environment (the 'parahippocampal place area' or PPA; Epstein & Kanwisher, 1998). fMRI was used to identify these brain areas in subjects, and their activations were then measured when the subjects viewed superimposed photographs of houses and faces, as in Fig. 12. Despite this spatial superimposition, the activations of the FFA and PPA were highly dependent on which stimulus subjects attended to, indicating an object-based attentional modulation of these areas' activations. Furthermore, even task-irrelevant features of the attended stimulus resulted in activation of the corresponding neural areas. When subjects had to attend only to a small motion in the face, for instance, *both* the motion area (MT/MST) and the FFA were activated, though again in this case the PPA was not activated. Again, it seemed that entire objects were being selected, rather than individual features.

4.2. The priority of spatiotemporal features

This strong view of feature processing, wherein *all* the features of an object are necessarily encoded when an object is attended, breaks down at high attentional loads. In the MOT paradigm (see Section 2.5), for example, successfully attending to the targets throughout the tracking phase appears to result in the encoding of spatiotemporal properties such as location and direction of motion, but *not* featural properties such as color and shape (Scholl et al., 2001b). To investigate whether items' locations were encoded as a result of being tracked, an item disappeared suddenly at the end of the motion phase, and subjects had to report the missing object's location using the mouse, and also indicate whether that item was a target or a distractor. As expected, performance was much better for successfully tracked target items compared to the unattended distractors. Similar results held when subjects had to use the mouse after the MOT motion phase to indicate the direction in which an item

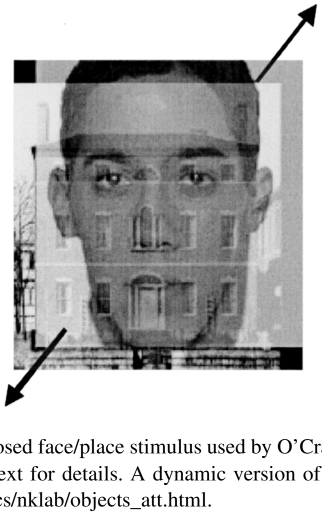

Fig. 12. Example of a superimposed face/place stimulus used by O'Craven et al. (1999) in an fMRI study of object-based attention. See text for details. A dynamic version of the figure is also available on the Internet at http://web.mit.edu/bcs/nklab/objects_att.html.

had been moving. A striking dissociation emerged, however, when certain other properties were examined. To investigate whether items' colors and shapes are encoded as a result of being tracked in MOT, the colors and shapes of items were occasionally permuted when the items were occluded or 'flashed', and subjects reported these properties when a single probe item disoccluded or 'unflashed' as a simple placeholder. (After the disocclusion of the placeholder, for instance, the subject would have to report what color it had been a moment before, and whether or not it was a target.) Here, surprisingly, property encoding was very poor, and was no better for attended targets than for unattended distractors: in the context of the attentional load induced by MOT, items' spatiotemporal properties but not their featural properties seem to be reliably encoded as a result of attention.

This pattern of results is in line with other earlier results obtained using briefly presented static stimuli (e.g. Sagi & Julesz, 1985), and suggests that attended object representations may in some circumstances be represented more robustly in terms of their spatiotemporal properties (especially location) than their featural properties (see also Jiang, Olson, & Chun, 2000; Johnston & Pashler, 1990; Nissen, 1985; Quinlan, 1998; Simons, 1996). This view is also supported by anecdotal evidence from Balint syndrome, where the objects seen by simultanagnosic patients do not seem to be tied to any particular set of visual features enjoyed by that object. Such patients, for instance, will often see many of the colors from each of the objects in the display 'float' through the single object which they are perceiving (Robertson, Treisman, Friedman-Hill, & Grabowecky, 1997; see also Humphreys, Cinel, Wolfe, Olson, & Klempen, 2000). As a whole, the evidence relating objects and features suggests that object-based processing may often trump feature-based processing, but that not all features are created equal: in some circumstances, spatiotemporal features may be more tightly coupled with object representations than are surface-based features such as color and shape.

5. Dynamic objects in space and time

The majority of the studies discussed in previous sections were concerned with demonstrating that attention can select discrete objects. Having established that objects can be units of attention, we can also ask about the dynamic nature of object representations. Two such issues are explored in this section: the maintenance of attended object tokens over time, and the possibility of attending to simple motions and events.

5.1. Maintaining object representations through time and motion

When attended objects move about the visual field, what factors constrain sustained attentional allocation to those items? An example of research addressing this question is discussed below, as are two general theories of how object tokens are maintained and updated.

5.1.1. Tracking items through occlusion

The MOT paradigm (Section 2.5) is well-suited to studying questions about the maintenance of object tokens, since only such sustained attention over time can distinguish the otherwise identical targets and distractors after the motion begins. This task has been used, for instance, to determine whether attended object representations survive interruptions in visibility during their motion (Scholl & Pylyshyn, 1999). Subjects tracked four small squares in a field of eight squares in total in a display which contained occluders (which were either drawn on the screen, or else were invisible but still functionally present). Subjects were able to successfully track even when the items were briefly (but completely) occluded at various times during their motion, suggesting that occlusion is taken into account when computing enduring perceptual objecthood (see also Tipper et al., 1990; Yantis, 1995). Unimpaired performance in the context of these occluders, however, required the presence of accretion and deletion cues along fixed contours at the occluding boundaries. Performance was significantly impaired when items were present on the visual field at the same times and to the same degrees as in the occlusion conditions, but disappeared and reappeared in ways which did not implicate the presence of occluding surfaces, for example by imploding and exploding into and out of existence instead of accreting and deleting along a fixed contour (see Fig. 13 for a schematic depiction of these conditions).

This pattern of results confirms that the circuits responsible for the 'attentional pursuit' of the items in this task are not simply robust in the face of any interruption in spatiotemporal continuity, but rather have a specific tolerance for interruptions consistent with occlusion. In other words, the local dynamics of items during brief disappearances help define what 'counts' as a dynamic visual object: items which disappear and reappear via accretion and deletion along a fixed contour are represented as persisting objects, and can be tracked in MOT, whereas those which disappear in other ways cannot be so tracked, as the disappearances seem to disrupt their continuing representation as the same object.

5.1.2. Object files

One general account of how object representations are maintained over time is the 'object file' theory (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992; Treisman, 1988, 1993; see also Kahneman & Henik, 1981). One traditional model of visual experience contends that visual stimuli are identified as objects when their visual projections activate semantic representations in long-term memory, and that visual experience consists of shifting patterns of this type of long-term memory activation. Kahneman et al. (1992) note the shortcomings of this view. It appears to be the case, for instance, that objects can be perceived and tracked even when they remain unidentified. Furthermore, when objects are initially mis-identified, and later correctly recognized, there is still never any doubt that the object involved was the same object. "Two identical red squares in successive fields may be perceived as distinct objects if the spatial/temporal gap between them cannot be bridged, but the transformation of frog into prince is seen as a change in a single visual object." (Kahneman et al., 1992, p. 179). Kahneman and Treisman argue that an intermediate level of representation is needed to mediate this latter task, which they call 'object files'.

On this theory, attending to an object in the visual field causes a temporary 'object file' representation to be created. Object files store information about the properties of visual objects (e.g. their colors, shapes, and current locations), but are allocated

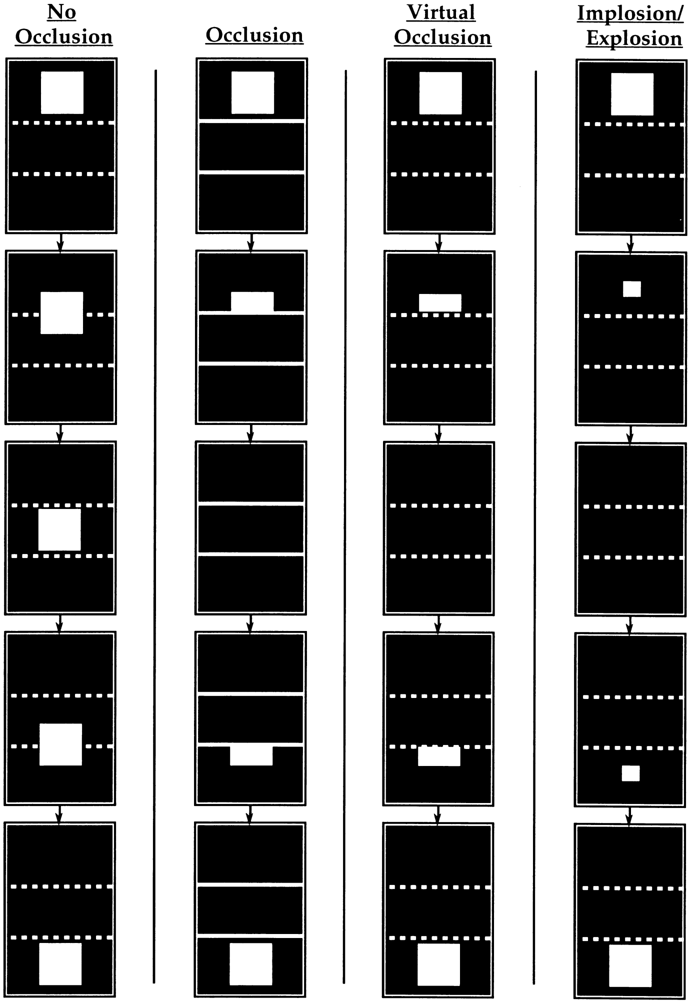

Fig. 13. Some of the different 'occlusion' conditions from Scholl and Pylyshyn (1999). The inherently dynamic nature of the occlusion conditions makes them difficult to represent in a static medium, but here they are presented as sequences of static 'snapshot' diagrams. In each condition, an item travels downward throughout five sequential frames of motion, interacting with a hypothetical occluder position (not to scale). Solid occluder boundaries represent visible occluders, while dashed occluder boundaries represent invisible occluders (presented to aid comprehension). See text for details. Animations of these conditions are available for viewing over the Internet at http://pantheon.yale.edu/~bs265/bjs-demos.html (Adapted from Scholl and Pylyshyn (1999).)

and maintained primarily on the basis of spatiotemporal factors. Three operations are involved in managing object files: (a) a *correspondence* operation, which determines for each object whether it is novel, or whether it moved from a previous location; (b) a *reviewing* operation, which retrieves an object's previous characteristics, some of which may no longer be visible; and finally (c) an *impletion* operation, which uses both current and reviewed information to construct a phenomenal percept, perhaps of object motion. When the features of two objects at different times match (via the reviewing process), the correspondence operation is thought to be facilitated, and the two objects are seen as temporal stages of a single enduring object in the world. When the features do not match, however, the correspondence between those items is inhibited.

This idea was tested with the 'object reviewing' paradigm (Kahneman et al., 1992). A single trial in this paradigm consists of two successive displays, as in Fig. 14. Each display contains small boxes, each of which may contain a single letter, and various manipulations are employed so that particular boxes in the first display are seen as continuous with particular boxes in the second display. The first ('preview') display contains two or more letters-in-boxes, while the second ('target') display contains a single letter-in-a-box, which can either match the letter from 'that' box in the initial display, can contain the letter from a 'different' box from the original display, or can contain an entirely novel letter. Subjects must simply name the single letter in the second display, and the typical result is that such response times are faster when that target is the same letter that filled the corresponding box in the first display (see Fig. 14). Kahneman et al. (1992) call this type of priming the 'object-specific preview effect': a preview facilitates or inhibits the processing of a target only if the preview and target are seen as states of the same object.

5.1.3. Visual indexing

A related theory called 'visual indexing' (Pylyshyn, 1989, 1994, 2001) complements the object file framework by postulating a mechanism whereby object-based individuation, tracking, and access are realized. In order to detect even simple geometrical properties among the elements of a visual scene (e.g. being collinear, or being 'inside'), Pylyshyn argues that the visual system must be able to simultaneously reference – or 'index' – multiple objects. This need is met in Pylyshyn's model by 'visual indexes', which are independently assigned to various items in the visual field on the basis of bottom-up salience cues, and which serve as a means of access to those items for the higher-level processes that allocate focal attention. In this regard, they function rather like pointers in a computer data structure: they reference certain items in the visual field (identifying them as distinct objects), without themselves encoding any properties of those objects. These indexes were referred to in early work as 'FINSTs', for FINgers of INSTantiation, due to the fact that physical fingers work in an analogous way: they can individuate and track items, and provide a means to determine relations such as 'to the left of', but they cannot by themselves encode an object's color or global shape. Visual indexes are thought to be assigned to objects in the visual field regardless of their spatial contiguity (in contrast with spotlight models), but with the restriction that the architecture of the

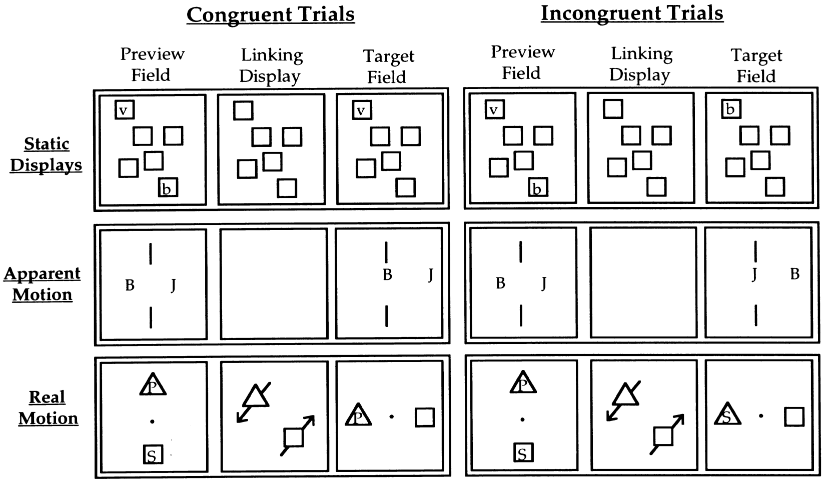

Fig. 14. Displays used by Kahneman et al. (1992). In the static displays, the target is seen as the same object as one of the previews, because it appears in the same location. In the apparent motion displays, the target (i.e. the stimulus between the two lines) is seen as the same object as one of the previews, because it is seen to arrive at its location via apparent motion from one of the preview locations. The same holds for the real-motion displays, mutatis mutandis. In each case, congruent information facilitates target naming, while incongruent information hampers performance. In control conditions (not pictured) in which the target is not seen as the same object as one of the previews, no such effects are observed. (Adapted from various figures in Kahneman et al. (1992).)

visual system provides only about *four* indexes. Furthermore, the indexes are sticky: if an indexed item in the visual field moves, the index moves with it.

Pylyshyn (2001) describes several types of experiments illustrating the need for and the operation of such visual indexes. This paper also stresses the data-driven (or, in Pylyshyn's terms, 'preconceptual') nature of the operation of these visual indexes. This aspect of the proposal serves to link visual processing up with the world, providing an exit to the regress in which various representational systems are explained in terms of other representational systems. If a significant portion of the indexing process is truly data-driven, then this might be a mechanism which 'gets vision off the ground'. In this sense, the visual indexing theory is intended to be a sort of interface between the world and the mind, and could underlie higher-level types of object-based processing.[11]

[11] In particular, Pylyshyn argues that the MOT paradigm (see Section 2.5) is a multi-stage process, and that the actual tracking itself illustrates the operation of the visual indexing system. Indeed, Pylyshyn created this paradigm in order to test the indexing theory (Pylyshyn & Storm, 1988). In this paper, in contrast, I have treated MOT as simply involving a standard type of attention. This is because MOT enjoys the salient properties of our pretheoretic notion of attention (it is selective, capacity-limited, and effortful), while it is unclear what aspects of MOT suggest or support a lower-level interpretation. At present, there seems to be no evidence ruling out the hypothesis that paradigms like MOT involve a standard type of attentional selection and 'pursuit', which is simply allocated in a rather complex way, 'split' between multiple items.

Both the object file and visual indexing frameworks embody theoretical assumptions which have been useful for guiding research on object-based attention. Perhaps the most basic assumption of these theories is simply that a level of visual processing exists in which the visual field is parsed and tracked as distinct objects, which are nevertheless not analyzed or recognized as *particular* objects. This is reminiscent of the evidence presented in Section 4 that spatiotemporal properties are more tightly bound to object representations than are surface-based properties.

5.2. Attention, sprites, and event perception

Whereas nearly all of the work on attention reviewed above has concerned either static objects or objects which happened to be in motion, attention may also interact directly with information which is *inherently* dynamic, such as simple stereotypical motions. Cavanagh and colleagues raise this possibility in their contribution to this special issue (Cavanagh et al., 2001). They suggest that the stereotypical motions of familiar objects – such as a person walking or a hand waving – may be stored as such, and that these 'units' of motion may facilitate recognition of the events and the objects participating in them. These stored representations of simple motions, termed *sprites*, are accessed or modeled by attention when viewing dynamic scenes. Such 'animation' of stored stereotypical motions is hypothesized to be among the visual system's standard repertoire of visual routines (in the sense of Ullman, 1984). As with other 'chunking' data structures (e.g. schemas, scripts), attentional sprites let familiar objects and events be recognized even from very sparse dynamic information in the scene. A complex set of sprites, for instance, would be responsible for the robust perception of biological motion (e.g. a person walking) which can arise when viewing even a very simple 'point light walker' composed of around 11 points of light in motion (Johansson, 1973).

The experiments reported by Cavanagh et al. (2001) focus on the attentional demands of using sprites. In particular, they explore whether simple patterns of points in motion can be discriminated without attention. Two such discriminations are tested: pairs of points 'tumbling' or 'orbiting' around a center point (Fig. 15a), and simple biological motion in one of two directions. When such dynamic stimuli are used in visual search tasks, so that the subjects must quickly determine whether a target motion is present in a field of distractor motions, the time taken to make this decision varies with the set size, a result which is taken to indicate that attention is required to 'animate' the sprites used to discriminate even simple motion patterns.

Beyond simple motion patterns, it is also possible that certain inherently dynamic *events* may serve as 'objects' of attention. Consider, for example, the perception of causality in simple 'launch displays', wherein one item is seen to hit another (Fig. 15b). It has been argued that the perception of such events is mediated by automatic low-level processes (Michotte, 1946/1963; see Scholl & Tremoulet, 2000, for a review). Such stimuli – and others, such as pushes, pulls, and even chases – are perceived in *causal* terms which go beyond the objective kinematics of the items in the display. It is possible that attentional sprites of the sort introduced here by Cavanagh et al. (2001) play a role in mediating such percepts by 'animating' the

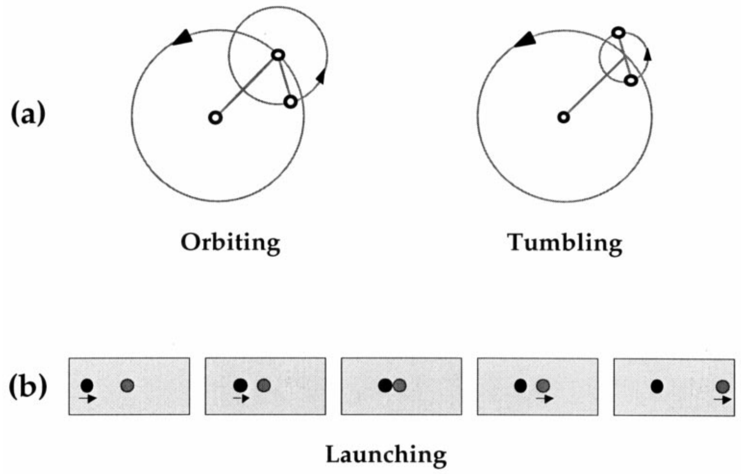

Fig. 15. Examples of simple events and motion patterns. (a) The 'tumbling' versus 'orbiting' pairs which had to be discriminated in visual search tasks in the experiments reported by Cavanagh et al. (2001). (Only the dots are actually drawn.) (b) A simple 'launch' event that has been thought to be perceived 'automatically' (Michotte, 1946/1963; Scholl & Tremoulet, 2000). Can such inherently dynamic animations and events serve as units of attention?

event schemas, and that in this way events might serve as units of attention. Such possibilities remain an intriguing focus for future research.

6. What is a visual object?

In previous sections of this paper, we have identified several different constraints on what can count as 'objects' of attention. For example, we have seen several instances of objects surviving both static and dynamic occlusion (Behrmann et al., 1998; Moore et al., 1998; Scholl & Pylyshyn, 1999; see also Tipper et al., 1990; Yantis, 1995), and we have seen that standard object-based effects are replicated with stimuli which we would intuitively characterize as groups, parts, and surfaces (see Section 3). Beyond these general categories, however, an important task for future work will be to determine the precise properties which mediate the degree to which visual feature clusters are treated as objects of attention. Some researchers, such as David Marr, have been pessimistic about the possibility of providing a useful answer to this question:

> Is a nose an object? Is a head one? Is it still one if it is attached to a body? What about a man on horseback? These questions show that the difficulties in trying to formulate what should be recovered as a region from an image are so great as to amount almost to philosophical problems. There is really no answer to them – all these things can be an object if you want to think of them that way, or they can be part of a larger object. (Marr, 1982, p. 270)

Marr's pessimism is certainly appropriate when considering the mind as a whole; certainly, for instance, we can *conceive* of almost anything as an object (for philosophical treatments, see Hirsch, 1982; Wiggins, 1980). With regard to what mental processes such as visual attention *treat* as objects, however, there may be well-defined answers to such questions.

The majority of work on object-based attention to date has been focused on demonstrating *that* object-based attention exists in various situations, independently of location-based and feature-based attention. In addition, some recent studies have begun to use the tools described in Section 2 to explore more directly what can count as an object of attention. Three such studies are briefly described here.

The first two studies employed the divided attention and spatial cueing paradigms which have previously revealed 'same-object advantages' (see Sections 2.3 and 2.4). In the original object-based cueing study, Egly, Driver, and Rafal (1994) observed a same-object advantage using pairs of rectangles as objects (see Fig. 3a). Avrahami (1999) set out to determine which features of these rectangles were crucial for the effect. She found that their closure was neither necessary nor sufficient: a same-object advantage was observed with simple sets of parallel lines, but not with certain distorted versions of the enclosed rectangles. Other researchers have similarly explored same-object advantages in the divided attention paradigm (Watson & Kramer, 1999). Subjects in these experiments viewed pairs of 'wrenches' as stimuli, and had to decide from extremely brief (50 ms) presentations whether the pair of wrenches contained both an open-ended wrench and a bent-end wrench. When these two features were on the same visual object, they reasoned, this response should be speeded. Using stimuli such as those in Fig. 16 (which shows 'same-object' trials for each condition), Watson and Kramer demonstrated that objecthood can be defined by uniformly connected regions such as those in Fig. 16a, but not by non-uniformly connected regions such as those in Fig. 16b (see also Kramer & Watson, 1996; Van Lier & Wagemans, 1998). They also showed that the magnitude of the same-object effect was attenuated when the ends of the wrenches were easily differentiable from the shafts by concave cusps as in Fig. 16c, compared to when such cusps were either non-existent (Fig. 16d) or not as pronounced (Fig. 16a) (see also Driver & Baylis, 1995; Hoffman & Singh, 1997). A final conclusion from this research was that the nature of visual objecthood varied by task (see also Brawn & Snowden, 2000; Lamy & Tsal, 2000): in other tasks, the existence of these cusps did not make a difference, and even regions which were not uniformly connected were treated as objects.

The issue of what can count as an object of attention is also addressed by Scholl et al. (2001a) in a MOT experiment involving a technique they call 'target merging'. Though subjects still attempted to track multiple independently and unpredictably moving items, the nature of these items was altered so that target/distractor pairs were perceived as single objects – with a target at one end and a distractor at the other end. For example, the pair might be drawn as a simple line segment connecting the two points, or as the convex hull of the two items. Each of the diagrams in Fig. 17, for instance, represents two targets and two distractors paired in various ways. (All of the actual experiments involved eight items in total paired into four target/distractor pairs.) Crucially, each 'end' of a pair still moved completely indepen-

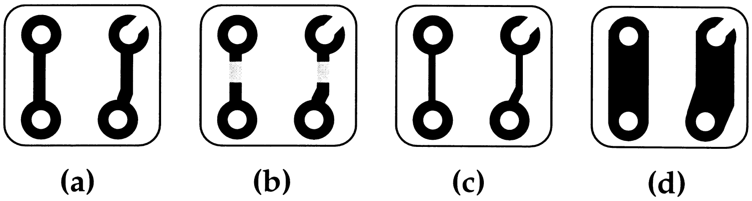

Fig. 16. Depictions of some of the 'wrench' stimuli used by Watson and Kramer (1999). In all displays, subjects must determine from brief (50 ms) presentations whether the display contains both a bent-end wrench and an open-ended wrench. All displays shown here represent 'same-object' target trials, since the two crucial features are always depicted on the same wrench. (a) A uniformly-connected display; (b) a non-uniformly connected display; (c) a display with salient cusps at each end of each wrench; (d) a display with no cusps at the end of the wrenches. See text for details. (Adapted from Watson and Kramer (1999).)

dently. Using a between-subjects design and identical sets of trajectories and target selections for each condition, Scholl et al. (2001a) find that tracking performance is radically impaired when the to-be-tracked items are undifferentiated parts or ends of larger objects such as lines (Fig. 17b) or 'rubber bands' (Fig. 17c). (In these cases, object-based attention is palpable: observers can feel attention being pulled in to the entire line or rubber band as the tracking phase unfolds.) This method is then used to explore the roles of part structure, connectedness, and other forms of perceptual grouping on visual objecthood. For instance, when subjects had to track ends of 'dumbbells' as in Fig. 17d, performance was worse than with boxes alone (Fig. 17a), but better than with lines alone (Fig. 17b) – presumably because of the salient parts at each end. In more complex cases, the precise nature of the connections seemed crucial: for instance, tracking was greatly impaired with 'Necker Cubes' (Fig. 17e), but not in the similar control condition depicted in Fig. 17f.

In most cases, given enough time and leisure, we are free to consider almost anything as an object. As these experiments demonstrate, however, objecthood is more well-defined at earlier levels of visual analysis. To get at such earlier levels, most investigators (e.g. Watson & Kramer, 1999) have used briefly presented and often masked stimuli along with speeded responses. This manipulation confines processing to early mechanisms because of a temporal limitation: the displays are gone before higher levels of analysis have a chance to come into play. The advantages of this method come with a cost, however: they result in small and imperceptible effects. In MOT, in contrast, the higher-level processes are constrained not by temporal limits but by overall sustained attentional load. It is trivially easy to track a *single* end of a line using focal attention, but the higher-level processes which make this possible are not available when attentional capacity is strained by the high load induced by MOT: in this latter case only a limited class of 'visual objects' can be tracked. The experiments described in this section report very preliminary results concerning the nature of visual objecthood, but hopefully these methods can continue to be used in the future to comprehensively explore the properties which give rise to objects of attention.

7. Beyond vision science

To this point, the discussion of objects and attention has been confined largely to aspects of visual processing in adults. In fact, however, the relation between objects and attention is also a central concern in the study of other modalities and even other sub-fields of cognitive science. Since a major goal of this special issue as a whole is to explore such connections, this penultimate section will address two such areas: *auditory* objects of attention (the topic of the paper by Kubovy & Van Valkenburg, 2001) and the infant's object concept (the topic of the paper by Carey & Xu, 2001).

7.1. Auditory objects of attention

There are many analogies between phenomena of object-based visual attention and phenomena of grouping and 'streaming' in audition. Albert Bregman (1990), in his influential book *Auditory scene analysis: the perceptual organization of sound*, pioneered a theory in which auditory scenes are grouped into and perceived as distinct auditory streams or objects of audition: "The stream plays the same role in auditory mental experience as the object does in visual." (p. 11). Each stream is perceived as containing those parts of the incoming auditory scene which 'go together'. In most cases, such analysis is tremendously useful since these different streams will emanate from different sources in the environment.

Kubovy and Van Valkenburg (2001) provide an overview of how to best conceptualize auditory objects, and relate them to visual objects. Early theories, they note,

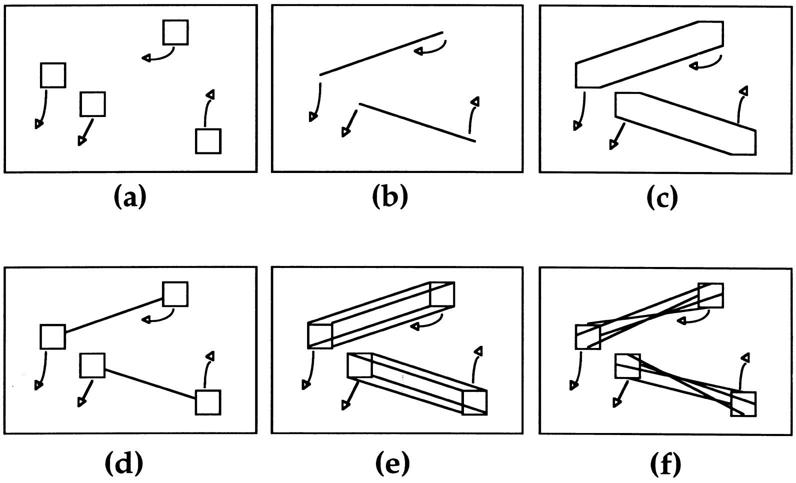

Fig. 17. Sample *target merging* displays from the MOT tasks of Scholl et al. (2001a). Each display shows four items, each of which always moves independently from all other items. (Actual displays had eight items in total.) In displays (b) through (f), the items are merged into pairs in various ways, with each pair always consisting of a target and a distractor, and subjects must track one *end* of each pair. Such manipulations greatly affect tracking performance. See text for details. Animations of these conditions are available for viewing over the Internet at http://pantheon.yale.edu/~bs265/bjs-demos.html.

tended to map auditory spatial processing (i.e. the computation of a sound source's location) onto visual spatial processing, and to map auditory frequency onto a visual property such as color. In contrast to this 'spatial' mapping, Kubovy and Van Valkenburg suggest that a more useful mapping is between auditory frequency and visual space. Just as visual objects exist in space-time, so auditory objects exist in pitch-time. This mapping is motivated by Kubovy's theory of 'indispensable attributes' (Kubovy, 1981), which notes that spatial separation is a necessary precondition for numerosity judgments in vision, while separation in frequency space is a necessary precondition for numerosity judgments in audition. (Two tones of different pitches coming from the same location will be judged as distinct, while two tones of the same pitch coming from different locations will be heard as a single sound.) Auditory spatial processing, in contrast, is seen as supporting visual spatial processing in the service of action, rather than being involved in object formation. Kubovy and Van Valkenburg summarize the existing evidence for this view, as well as the evidence that auditory processing contains separate 'what' and 'where' streams, as does vision. In addition, they discuss how attention interacts with each stream, proposing that attention is typically drawn only to indispensable attributes in each modality (i.e. primarily to objects and locations but not colors in vision, and primarily to pitches but not spatial locations in audition).

The view of objecthood which emerges from this theory is intended to be cross-modal. Early perception – both auditory and visual – aggregates 'elements', which then undergo grouping to form various perceptual organizations. Each of these perceptual organizations is then a potential 'object', and actual objects are formed via attentional selection, which results in figure-ground segregation. An advantage of this particular conception of objecthood is that it can be stated in this modality-independent way. Note, though, that the choice of which of these levels count as the 'objects' is somewhat arbitrary. Most of the 'objects' (and sometimes 'proto-objects'; see Section 3.6) that have figured in the object-based attention work might be termed 'elements' on this model. Of course, it remains an important topic for future research to determine the properties which mediate unit formation at each of these levels in Kubovy and Van Valkenburg's theory: the early 'elements', the mid-level 'groups', and the final 'objects' which emerge as a result of attention (see also Section 3.6).

In addition to these general lessons on how to think about auditory objects and relate them to visual objects, it is also possible to draw more specific analogies between auditory phenomena and experiments on object-based visual attention. Two such examples are discussed in the remainder of this section. First, recall the evidence (presented in Section 2) that attention in some contexts is not a simple unitary spotlight. In the MOT task, for instance, attention can be split between multiple items in space, rather than being spread between them. Keeping to the analogy between visual space and auditory frequency, similar results are obtained: just as some of the results described in Section 2 seem inconsistent with attention to a single region of visual space, so do the results of some audition experiments seem inconsistent with attention to a single region of frequency space. For example, listeners are able to simultaneously monitor for both a low- and a high-frequency tone just as easily as they can monitor for the two tones in sequential intervals (Johnson & Hafter, 1980). A more direct source of evidence comes from a monitoring situation where subjects have to determine which of two sequential temporal intervals contains a target tone. In this situation, when the target tones are usually at two separated frequencies, subjects perform well at both frequencies, but perform poorly for those targets which unexpectedly have tones between these typical frequencies (Macmillan & Schwartz, 1975). This demonstrates that attention can be split between multiple auditory tones rather than simply spread between them in frequency space.

A second example of convergences between the general principles of visual and auditory processing involves the processing of occlusion. Many researchers of auditory attention (e.g. Dannenbring, 1976; Warren, 1982) have studied how sounds moving in frequency space can seem to continue 'behind' auditory occluders, such as sudden bursts of noise (see Fig. 18a). This type of situation is in some ways analogous to the experiments of Scholl and Pylyshyn (1999) discussed in Section 5.1.1, where spatial movement is analogous to movement in frequency space. In those experiments the nature of the local disappearance at the occluding boundary made a crucial difference to whether the item could be tracked through that boundary: when the items 'imploded' and 'exploded' at the occluding boundaries, for example, performance was severely impaired (see Fig. 13). Auditory researchers have observed similar effects. For instance, if the initial frequency ends (Fig. 18b) – or begins to gradually diminish in amplitude (Fig. 18b) – a moment before the burst of noise, then the auditory percept of continuation is severely reduced or eliminated (Bregman, 1990; Bregman & Dannenbring, 1977; Warren et al., 1972). In both cases, continuity through occlusion occurs only when all of the 'disappearing' of the tracked visual or auditory object occurs along the contour of the occluding boundary. Bregman (1990) identifies the general principle involved here: "The perceptual systems, both visual and auditory, must use a very accurate analysis of the structure of the sensory evidence to determine whether the parts separated by the occluding material show sufficient agreement with one another to be considered parts of the same thing or event." (p. 347). Such analogies are provocative, if

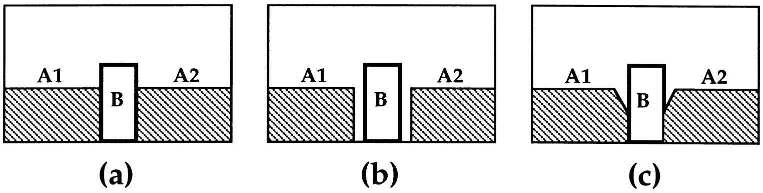

Fig. 18. Depictions of auditory events used by Bregman (1990), Bregman and Dannenbring (1977), and Warren, Obusek, and Ackroff (1972). In each diagram, the horizontal axis represents time, while the vertical axis represents intensity. (a) Sound A1 ceases just as sound B begins, and sound A2 begins just as sound B ceases. Subjects perceive A1 continuing behind the 'auditory occluder' of B. This continuity percept is attenuated or destroyed, however, when A1 stops (b) or diminishes in intensity (c) just before the onset of B. See text for details. (Adapted from Bregman (1990).)

only for the heuristic value they bring to each field in terms of generating experiments and theories, and they clearly merit further study.

7.2. Object-based attention and the infant's object concept

Cognitive developmental psychology is another area of study which has often focused on issues of objects and attention. Using looking-time measures to study the infant's 'object concept', developmental psychologists have demonstrated that infants even a few months old have a substantial amount of 'initial knowledge' about objects in domains such as physics and arithmetic (for recent reviews and overviews, see Baillargeon, 1995; Carey, 1995; Spelke, 1994; Spelke et al., 1995; Wynn, 1998). Traditional discussions of the nature of such 'initial knowledge' have assumed an implicit dichotomy between 'perception' and 'cognition' (e.g. Bogartz, Shinskey, & Speaker, 1997; Kellman, 1988; Leslie, 1988; Spelke, 1988a,b), and from within this dichotomy 'perception' was often found wanting, largely because it was thought not to be object-based (see the quote from Spelke, 1988b, in Section 1.1). Since 'perception' was thought not to traffic in discrete objects, but 'thought' was, the correct explanations for the infancy experiments were assumed to be 'conceptual' in nature (see Scholl & Leslie, 1999, for discussion). Of course, all of the evidence discussed in this article belies this characterization of perception, and if parts of perception can indeed be object-based, then it is possible that mechanisms of object-based attention play an important role in explaining these infancy results.

Scholl and Leslie (1999) drew just this conclusion, and identified several convergences between these two fields (see also Leslie, Xu, Tremoulet, & Scholl, 1998). To take one example, recall the priority for spatiotemporal over featural properties that was found by Scholl et al. (2001b) (see Section 4.2), and which is inherent in theories of object files and visual indexing (Section 5.1). This pattern mirrors the maturational differences in property encoding obtained with 10–12-month-old infants by Xu and Carey (1996). Ten-month-old infants, for instance, will use spatiotemporal information (seeing two unconnected items emerge from behind a screen at the same time) but not featural information (seeing a red item and a green item emerge sequentially) to infer the existence of two distinct objects behind the screen, as in Figs. 19 and 20 (Xu & Carey, 1996). Twelve-month-olds, in contrast, will use both sorts of information, like adults. In a similar vein, 4-month-old infants have been shown to use spatiotemporal information to infer that two parts are in fact a single unitary object (e.g. the fact that two parts separated by an occluder undergo common motion) but not featural information (e.g. the fact that two stationary parts separated by an occluder have similar colors and/or shapes; Kellman & Spelke, 1983; Kestenbaum, Termine, & Spelke, 1987; see also Van de Walle & Spelke, 1996). (Again, adults and older children will use both sorts of information.) Furthermore, infants at these ages appear to use *only* spatiotemporal information to assess an object's unity: in the situation described above, for example, infants are perfectly happy to conceive of the two parts in common motion as a single object, despite the

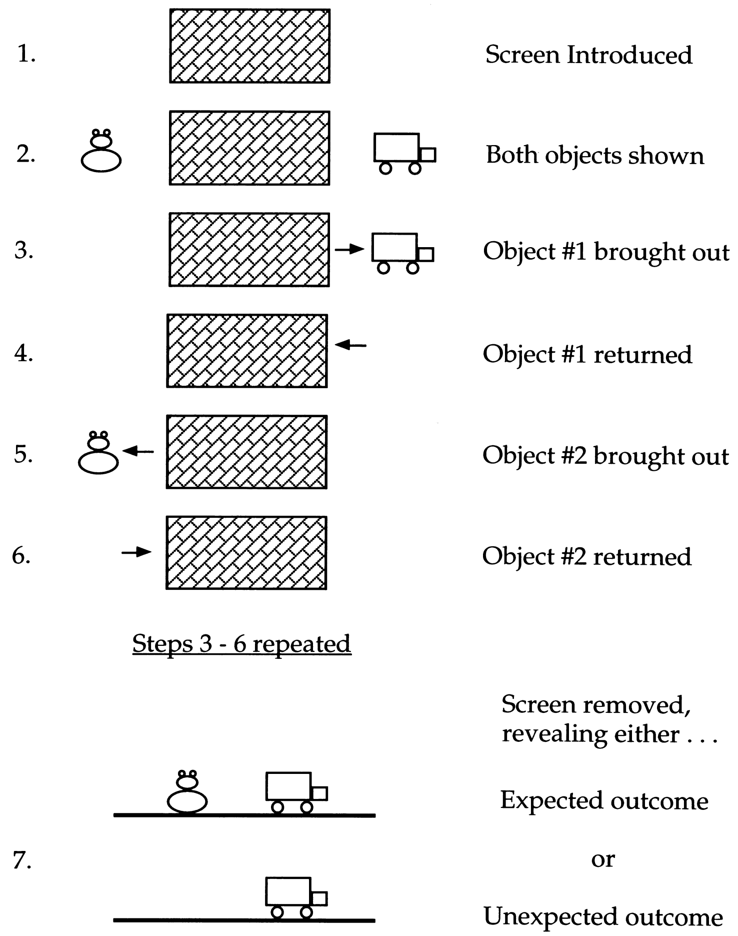

Fig. 19. The ‘spatial’ condition from Xu and Carey (1996). See text for details. (Adapted from Xu and Carey (1996).)

fact that their colors and shapes suggest strongly (to older children and adults) that they are distinct objects (but cf. Needham, 1997).

Carey and Xu (2001) address the relationship between object-based visual attention in adults and this type of infancy work at length, identifying several other convergences, as well as their limitations. They suggest that the adult mind has two primary representational systems for individuating objects. One, which they call the ‘mid-level object file system’, involves the types of attentional processes discussed throughout this article. The second is a ‘kind-based’ system, which is fully conceptual, and can often override the attention-based tracking system (e.g. when you decide that the computer on your desk is the same one that was there yesterday, despite the fact that you did not directly observe the spatiotemporal continuity

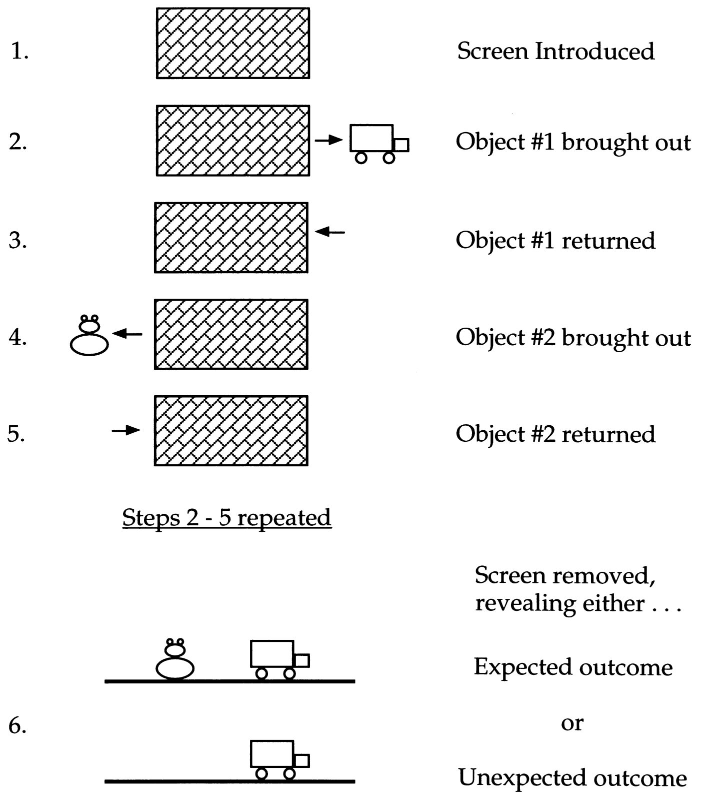

Fig. 20. The ‘property’ condition from Xu and Carey (1996). See text for details. (Adapted from Xu and Carey (1996).)

between the two). Whereas properties such as color and shape may play only a peripheral role in mid-level object tracking, they may play a more central role in kind-based individuation. Because young infants do not yet employ kind-based individuation, Carey and Xu stress that they are ideal ‘tools’ with which to study mid-level object-tracking: adults, in contrast, are ‘contaminated’ by both systems, which are often hard to distinguish. In addition to stressing the role of attention-based systems in infant cognition, Carey and Xu also identify several ways in which results from the infant cognition literature might usefully inform work on object-based attention. In particular, they describe a nuanced view of the relation between object-based attention and higher-level processes. They note that the representations formed by processes of object-based attention may still be conceptual in nature, despite their perceptual origin: for example, these representations may end up playing a prominent role in guiding further inferences and actions.

These connections between developmental work and work from vision science on adults should be an exciting topic for future research. Research in these two fields has until this point proceeded largely independently, and it seems certain that each field will have many heuristic insights to offer the other, and that in some cases researchers from the two fields may be studying the same mechanisms of object-based attention.

8. Conclusions: a case study in cognitive science?

The study of objects and attention is important, in the first instance, for intrinsic reasons: a fundamental task in the study of visual attention is to determine the nature of the basic units over which attention operates. As has been reviewed in this paper, the units of attention are often various kinds of visual *objects*. That this is true seems undeniable in the face of converging evidence from so many psychophysical and neuropsychological experiments. Still, there are many important questions which remain to be answered about object-based attention. These include the question of precisely which stimulus features define 'objecthood' from the perspective of the visual system (Section 6), and how (or if) object-based attention differs from notions of group-, surface-, part-, or event-based attention (Sections 3 and 5).

As the additional analogies with other fields (Section 7) begin to suggest, however, the study of objects and attention may also be of interest more generally. The evidence reviewed above consisted largely of experimental psychology and neuropsychology, but there has also been valuable recent input from computational modeling (e.g. Behrmann et al., 1998; Mozer, 1999) which was not reviewed here. Furthermore, in addition to the work on audition and cognitive development discussed in Section 7, the relation of objects and attention has also been of much interest to philosophers (e.g. Hirsch, 1997; Wiggins, 1997; cf. Xu, 1997) and even language researchers (e.g. Landau & Jackendoff, 1993). In few other areas of this young field have so many areas of study converged on so many similar ideas, and as such the research on this topic might be viewed as an emerging 'case study' in cognitive science.

Acknowledgements

B.J.S. was supported by NIH #F32-MH12483-01. For helpful conversation and/or comments on earlier drafts, I thank Jacob Feldman, Jerry Fodor, Steve Franconeri, Alex Holcombe, Alan Leslie, Ken Nakayama, Mary Peterson, Dan Simons, Anne Treisman, and Steve Yantis. Special thanks to Jon Driver and Zenon Pylyshyn for especially careful and critical readings of an earlier draft. Special thanks are also due to the members of my 1999 undergraduate seminar at Rutgers, 'Objects: A Case Study in Cognitive Science', for extremely helpful extended conversation.

References

Atchley, P., & Kramer, A. (in press). Object-based attentional selection in three-dimensional space. *Visual Cognition*, *8*(1).

Avrahami, J. (1999). Objects of attention, objects of perception. *Perception & Psychophysics*, *61*, 1604–1612.

Baillargeon, R. (1995). Physical reasoning in infancy. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 181–204). Cambridge, MA: MIT Press.

Baylis, G. (1994). Visual attention and objects: two-object cost with equal convexity. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 208–212.

Baylis, G., & Driver, J. (1993). Visual attention and objects: evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 451–470.

Baylis, G., & Driver, J. (1995). Obligatory edge assignment in vision: the role of figure and part segmentation in symmetry selection. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1323–1342.

Behrmann, M., & Tipper, S. (1994). Object-based visual attention: evidence from unilateral neglect. In C. Umilta, & M. Moscovitch (Eds.), *Attention and performance. Conscious and nonconscious processing and cognitive functioning* (Vol. 15, pp. 351–375). Cambridge, MA: MIT Press.

Behrmann, M., & Tipper, S. (1999). Attention accesses multiple reference frames: evidence from unilateral neglect. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 83–101.

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.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

Bogartz, R., Shinskey, J., & Speaker, C. (1997). Interpreting infant looking: the event set × event set design. *Developmental Psychology*, *33*, 408–422.

Boutsen, L., & Humphreys, G. (2000). Axis-based grouping reduces visual extinction. *Neuropsychologia*, *38*, 896–905.

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.

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.

Bregman, A. S., & Dannenbring, G. L. (1977). Auditory continuity and amplitude edges. *Canadian Journal of Psychology*, *31*, 151–159.

Caramazza, A., & Hillis, A. (1990). Levels of representations, co-ordinate frames, and unilateral neglect. *Cognitive Neuropsychology*, *7*, 391–445.

Carey, S. (1995). Continuity and discontinuity in cognitive development. In E. Smith, & D. Osherson (Eds.), *Thinking: Vol. 3. An invitation to cognitive science* (2nd ed., pp. 101–130). Cambridge, MA: MIT Press.

Carey, S., & Xu, F. (2001). Infant knowledge of objects: beyond object files and object tracking. *Cognition*, this issue, *80*, 179–213.

Cavanagh, P., Labianca, A., & Thornton, I. (2001). Attention-based visual routines: sprites. *Cognition*, this issue, *80*, 47–60.

Cave, K., & Bichot, N. (1999). Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin & Review*, *6*, 204–223.

Chen, Z. (1998). Switching attention within and between objects: the role of subjective organization. *Canadian Journal of Experimental Psychology*, *52*, 7–16.

Chun, M. (1997). Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 738–755.

Cooper, A., & Humphreys, G. (1999). *A new, object-based visual illusion*. Poster presented at the annual meeting of the Psychonomic Society, Los Angeles, CA.

Coslett, H. B., & Saffran, E. (1991). Simultanagnosia: to see but not two see. *Brain*, *113*, 1523–1545.

Culham, J. C., Brandt, S., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*, 2657–2670.

Culham, J. C., Cavanagh, P., & Kanwisher, N. (2001). Attention response functions of the human brain measured with fMRI. Manuscript submitted for publication.

Dannenbring, G. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology*, *30*, 99–114.

Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature*, *371*, 791–793.

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.

Downing, C., & Pinker, S. (1985). The spatial structure of visual attention. In M. Posner, & O. S. M. Marin (Eds.), *Attention and performance* (Vol. XI, pp. 171–187). London: Erlbaum.

Driver, J. (1998). The neuropsychology of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 297–340). Hove: Psychology Press.

Driver, J., & Baylis, G. (1995). One-sided edge assignment in vision: II. Part decomposition, shape description, and attention to objects. *Current Directions in Psychological Science*, *4*, 201–206.

Driver, J., & Baylis, G. (1996). Figure-ground segmentation and edge assignment in short-term visual matching. *Cognitive Psychology*, *31*, 248–306.

Driver, J., & Baylis, G. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The attentive brain* (pp. 299–325). Cambridge, MA: MIT Press.

Driver, J., Baylis, G., Goodrich, S., & Rafal, R. (1994). Axis-based neglect of visual shapes. *Neuropsychologia*, *32*, 1353–1365.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention, and phenomenal visual objects. *Cognition*, this issue, *80*, 61–95.

Driver, J., & Halligan, P. (1991). Can visual neglect operate in object-centered coordinates? An affirmative single case study. *Cognitive Neuropsychology*, *8*, 475–494.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*, 272–300.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.

Duncan, J. (1993a). Coordination of what and where in visual attention. *Perception*, *22*, 1261–1270.

Duncan, J. (1993b). Similarity between concurrent visual discriminations: dimensions and objects. *Perception & Psychophysics*, *54*, 425–430.

Duncan, J., & Nimmo-Smith, I. (1996). Objects and attributes in divided attention: surface and boundary systems. *Perception & Psychophysics*, *58*, 1076–1084.

Egly, R., Driver, J., & Rafal, R. (1994). Shifting visual attention between objects and locations: evidence for normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161–177.

Egly, R., Rafal, R., Driver, J., & Starrveveld, Y. (1994). Covert orienting in the split brain reveals hemispheric specialization for object-based attention. *Psychological Science*, *5*, 380–383.

Enns, J., & Rensink, R. (1998). Early completion of occluded objects. *Vision Research*, *38*, 2489–2505.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the visual identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*, 201–204.

Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, *14*, 155–160.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception & Psychophysics*, *40*, 225–240.

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 583–597.

Feldman, J. (1999). The role of objects in perceptual grouping. *Acta Psychologica*, *102* (1), 137–163.

Gibson, B. (1994). Visual attention and objects: one versus two, or convex versus concave? *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 203–207.

He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends in Cognitive Sciences*, *1*, 115–121.

He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in 3-D space. *Proceedings of the National Academy of Sciences USA*, *92*, 11155–11159.

Hirsch, E. (1982). *The concept of identity*. New York: Oxford University Press.

Hirsch, E. (1997). Basic objects: a reply to Xu. *Mind & Language*, *12*, 406–412.

Hochberg, J., & Peterson, M. A. (1987). Piecemeal perception and cognitive components in object perception: perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General*, *116*, 370–380.

Hoffman, D., & Richards, W. (1984). Parts of recognition. *Cognition*, *18*, 65–96.

Hoffman, D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *69*, 29–78.

Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics*, *30*, 283–290.

Holmes, G., & Horax, G. (1919). Disturbances of spatial orientation and visual attention, with loss of stereoscopic vision. *Archives of Neurology and Psychiatry*, *1*, 385–407.

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A., & Klempen, N. (2000). Fractionating the binding process: neuropsychological evidence distinguishing binding of form from binding of surface features. *Vision Research*, *40*, 1569–1696.

Humphreys, G. W., & Riddoch, M. J. (1993). Interactions between object and space systems revealed through neuropsychology. In D. Meyer, & S. Kornblum (Eds.), *Attention and performance* (Vol. XIV, pp. 183–218). Cambridge, MA: MIT Press.

Humphreys, G. W., & Riddoch, M. J. (1994). Attention to within-object and between-object spatial representations: multiple sites for visual selection. *Cognitive Neuropsychology*, *11*, 207–241.

Intriligator, J. M. (1997). *The spatial resolution of visual attention*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Irwin, D., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui, & J. McClelland (Eds.), *Attention and performance*, Vol. XVI. Cambridge, MA: MIT Press.

James, W. (1890). *The principles of psychology*. New York: Holt.

Jiang, Y., Olson, I., & Chun, M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 683–702.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201–211.

Johnson, D., & Hafter, E. (1980). Uncertain-frequency detection: cueing and condition of observation. *Perception & Psychophysics*, *28*, 143–149.

Johnston, J., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 843–856.

Johnston, W., & Dark, V. (1986). Selective attention. *Annual Review of Psychology*, *37*, 43–75.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy, & J. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, *24*, 174–219.

Kanwisher, N. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, *27*, 117–143.

Kanwisher, N. (1991). Repetition blindness and illusory conjunctions: errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 404–421.

Kanwisher, N., & Driver, J. (1992). Objects, attributes, and visual attention: which, what, and where. *Current Directions in Psychological Science*, *1*, 26–31.

Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for the perception of faces. *Journal of Neuroscience*, *17*, 4302–4311.

Kellman, P. (1988). Theories of perception and research in perceptual development. In A. Yonas (Ed.), *Perceptual development in infancy*: Vol. 20. *The Minnesota Symposium on Child Psychology* (pp. 267–281). Hillsdale, NJ: Erlbaum.

Kellman, P., & Spelke, E. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524.

Kestenbaum, R., Termine, N., & Spelke, E. S. (1987). Perception of objects and object boundaries by 3-month-old infants. *British Journal of Developmental Psychology*, *5*, 367–383.

Kramer, A., & Jacobson, A. (1991). Perceptual organization and focused attention: the role of objects and proximity in visual processing. *Perception & Psychophysics*, *50*, 267–284.

Kramer, A., & Watson, S. (1996). Object-based visual selection and the principle of uniform connectedness. In A. Kramer, M. Coles, & G. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 395–414). Washington, DC: APA Press.

Kramer, A., Weber, T., & Watson, S. (1997). Object-based attentional selection – grouped arrays or spatially invariant representations?: comment on Vecera and Farah (1994). *Journal of Experimental Psychology: General*, *126*, 3–13.

Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy, & J. Pomerantz (Eds.), *Perceptual organization* (pp. 55–99). Hillsdale, NJ: Erlbaum.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, this issue, *80*, 97–126.

LaBerge, D., & Brown, V. (1989). Theory of attentional operation in shape identification. *Psychological Review*, *96*, 101–124.

Laeng, B., Kosslyn, S., Caviness, V., & Bates, J. (1999). Can deficits in spatial indexing contribute to simultanagnosia? *Cognitive Neuropsychology*, *16*, 81–114.

Lamy, D. (2000). Object-based selection under focused attention: a failure to replicate. *Perception & Psychophysics*, *62*, 1272–1279.

Lamy, D., & Tsal, Y. (2000). Object features, object locations, and object files: which does selective attention activate and when? *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1387–1400.

Landau, B., & Jackendoff, R. (1993). ‘What’ and ‘where’ in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217–265.

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

Leslie, A. M. (1988). The necessity of illusion: perception and thought in infancy. In L. Weiskrantz (Ed.), *Thought without language* (pp. 185–210). Oxford: Oxford Science.

Leslie, A. M., Xu, F., Tremoulet, P., & Scholl, B. J. (1998). Indexing and the object concept: developing ‘what’ and ‘where’ systems. *Trends in Cognitive Sciences*, *2* (1), 10–18.

Luck, S., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Luria, A. R. (1959). Disorders of ‘simultaneous perception’ in a case of bilateral occipito-parietal brain injury. *Brain*, *83*, 437–449.

Macmillan, N., & Schwartz, M. (1975). A probe-signal investigation of uncertain-frequency detection. *Journal of the Optical Society of America*, *58*, 1051–1058.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics*, *58*, 977–991.

Marr, D. (1982). *Vision*. New York: W.H. Freeman.

Michotte, A. (1963). *The perception of causality* (T. Miles, & E. Miles, Trans.). New York: Basic Books. (Original work published 1946)

Moore, C., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: evidence from perceptual completion. *Psychological Science*, *9*, 104–110.

Most, S. B., Simons, D. J., Scholl, B. J., Jiminez, R., Clifford, E., & Chabris, C. F. (in press). How not to be seen: the contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological Science*.

Mounts, J., & Melara, R. (1999). Attentional selection of objects or features: evidence from a modified search task. *Perception & Psychophysics*, *61*, 322–341.

Mozer, M. (1999). Do visual attention and perception require multiple reference frames? Evidence from a computational model of unilateral neglect. *Proceedings of the 21st annual conference of the Cognitive Science Society* (pp. 456–461). Mahwah, NJ: Erlbaum.

Nakayama, K., He, Z., & Shimojo, S. (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In S. M. Kosslyn, & D. Osherson (Eds.), *Visual cognition*: Vol. 2. *An invitation to cognitive science* (2nd ed., pp. 1–70). Cambridge, MA: MIT Press.

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

Needham, A. (1997). Factors affecting infants' use of featural information in object segregation. *Current Directions in Psychological Science*, *6*, 26–33.

Neely, C., Dagenbach, D., Thompson, R., & Carr, T. (1998). *Object-based visual attention: the spread of attention within objects and the movement of attention between objects*. Paper presented at the 39th annual meeting of the Psychonomic Society, Dallas, TX.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1979). The control of information pickup in selective looking. In A. Pick (Ed.), *Perception and its development* (pp. 201–219). Hillsdale, NJ: Erlbaum.

Neisser, U., & Becklen, R. (1975). Selective looking: attending to visually specified events. *Cognitive Psychology*, *7*, 480–494.

Nissen, M. J. (1985). Accessing features and objects: is location special? In M. Posner, & O. Marin (Eds.), *Attention and performance* (Vol. XI, pp. 205–220). Hillsdale, NJ: Erlbaum.

O'Craven, K., Downing, P., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.

Palmer, S. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*, 441–474.

Palmer, S. (1999). *Vision science: photons to phenomenology*. Cambridge, MA: MIT Press.

Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Peterson, M. A., Gerhardstein, P., Mennemeier, M., & Rapcsak, S. (1998). Object-centered attentional biases and object recognition contributions to scene segmentation in left- and right-hemispheric-damaged patients. *Psychobiology*, *26*, 357–370.

Peterson, M. A., & Gibson, B. S. (1991). Directing spatial attention within an object: altering the functional equivalence of shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 170–182.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: a sketch of the FINST spatial index model. *Cognition*, *32*, 65–97.

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition*, *50*, 363–384.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, this issue, *80*, 127–158.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.

Quinlan, P. T. (1998). The recovery of identity and relative position from visual input: further evidence for the independence of processing of what and where. *Perception & Psychophysics*, *60*, 303–318.

Rafal, R. D. (1997). Balint syndrome. In T. Feinberg, & M. Farah (Eds.), *Behavioral neurology and neuropsychology* (pp. 337–356). New York: McGraw-Hill.

Rafal, R. D. (1998). Neglect. In R. Parasuraman (Ed.), *The attentive brain* (pp. 489–525). Cambridge, MA: MIT Press.

Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.

Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision Research*, *40*, 1469–1487.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, *8* (5), 368–373.

Reuter-Lorenz, P., Drain, M., & Hardy-Morais, C. (1996). Object-centered attentional biases in the intact brain. *Journal of Cognitive Neuroscience*, *8*, 540–550.

Robertson, I. & Marshall, J. (1993). *Unilateral neglect: clinical and experimental studies*. Hove: Erlbaum.

Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways: evidence from Balint syndrome. *Journal of Cognitive Neuroscience*, *9*, 295–317.

Sagi, D., & Julesz, B. (1985). What and where in vision. *Science*, *228*, 1217–1219.

Scholl, B. J., & Leslie, A. M. (1999). Explaining the infant's object concept: beyond the perception/cognition dichotomy. In E. Lepore, & Z. Pylyshyn (Eds.), *What is cognitive science?* (pp. 26–73). Oxford: Blackwell.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001a). What is a visual object? Evidence from target merging in multi-element tracking. *Cognition*, this issue, *80*, 159–177.

Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (2001b). The relationship between property-encoding and object-based attention: evidence from multiple object tracking. Manuscript submitted for publication.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4* (8), 299–309.

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

Simons, D. J. (1996). In sight, out of mind: when object representations fail. *Psychological Science*, *7*, 301–305.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1* (7), 261–267.

Singh, M., & Scholl, B. J. (2000). *Using attentional cueing to explore part structure*. Poster presented at the 2000 Pre-Psychonomies Object Perception and Memory meeting, New Orleans, LA.

Spelke, E. (1988a). The origins of physical knowledge. In L. Weiskrantz (Ed.), *Thought without language* (pp. 168–184). Oxford: Oxford Science.

Spelke, E. (1988b). Where perceiving ends and thinking begins: the apprehension of objects in infancy. In A. Yonas (Ed.), *Perceptual development in infancy* (pp. 197–234). Hillsdale, NJ: Erlbaum.

Spelke, E. (1994). Initial knowledge: six suggestions. *Cognition*, *50*, 431–445.

Spelke, E., Gutheil, G., & Van de Walle, G. (1995). The development of object perception. In S. Kosslyn, & D. Osherson (Eds.), *Visual cognition. An invitation to cognitive science* (2nd ed. pp. 297–330). Cambridge, MA: MIT Press.

Stuart, G., Maruff, P., & Currie, J. (1997). Object-based visual attention in luminance increment detection? *Neuropsychologia*, *35*, 843–853.

Styles, E. (1997). *The psychology of attention*. Hove: Psychology Press.

Subbiah, I., & Caramazza, A. (2000). Stimulus-centered neglect in reading and object-recognition. *Neurocase*, *6*, 13–31.

Tipper, S., & Behrmann, M. (1996). Object-centered not scene-based visual neglect. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1261–1278.

Tipper, S., Brehaut, J., & Driver, J. (1990). Selection of moving and static objects for the control of spatially directed action. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 492–504.

Tipper, S., Driver, J., & Weaver, B. (1991). Object-centered inhibition of return of visual attention. *Quarterly Journal of Experimental Psychology*, *43A*, 289–298.

Tipper, S., Jordan, H., & Weaver, B. (1999). Scene-based and object-centered inhibition of return: evidence for dual orienting mechanisms. *Perception & Psychophysics*, *61*, 50–60.

Treisman, A. (1988). Features and objects: the fourteenth Bartlett memorial lectures. *Quarterly Journal of Experimental Psychology*, *40*, 201–237.

Treisman, A. (1993). The perception of features and objects. In A. Baddeley, & L. Weiskrantz (Eds.), *Attention: selection, awareness, and control* (pp. 5–35). Oxford: Clarendon Press.

Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, *33*, 527–532.

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (1998). Transparent motion and object-based attention. *Cognition*, *66*, B13–B23.

Van de Walle, G., & Spelke, E. (1996). Spatiotemporal integration and object perception in infancy: perceiving unity vs. form. *Child Development*, *67*, 2621–2640.

Van Lier, R., & Wagemans, J. (1998). Effects of physical connectivity on the representational unity of multi-part configurations. *Cognition*, *69*, B1–B9.

Vecera, S. (1994). Grouped locations and object-based attention: comment on Egly, Driver, and Rafal (1994). *Journal of Experimental Psychology: General*, *123*, 316–320.

Vecera, S., Behrmann, M., & Filapek, J. (in press). Attending to the parts of a single object: part-based selection limitations. *Perception & Psychophysics*.

Vecera, S., Behrmann, M., & McGoldrick, J. (2000). Selective attention to the parts of an object. *Psychonomic Bulletin & Review*, *7*, 301–308.

Vecera, S., & Farah, M. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1–14.

Viswanathan, L., & Mingolla, E. (in press). Dynamics of attention in depth: evidence from multi-element tracking. *Perception*.

Ward, R., Goodrich, S., & Driver, J. (1994). Grouping reduces visual extinction: neuropsychological evidence for weight-linkage in visual selection. *Visual Cognition*, *1*, 101–129.

Warren, R. (1982). *Auditory perception: a new synthesis*. New York: Pergamon.

Warren, R., Obusek, C., & Ackroff, J. (1972). Auditory induction: perceptual synthesis of absent sounds. *Science*, *176*, 1149–1151.

Watson, D., & Humphreys, G. (1998). Visual marking of moving objects: a role for top-down feature-based inhibition in selection. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 946–962.

Watson, S., & Kramer, A. (1999). Object-based visual selective attention and perceptual organization. *Perception & Psychophysics*, *61*, 31–49.

Watt, R. J. (1988). *Visual processing: computational, psychophysical, and cognitive research*. Hillsdale, NJ: Erlbaum.

Wheeler, M., & Treisman, A. (1999). *Aspects of time, space, and binding in visual working memory for simple objects*. Paper presented at the 1999 Pre-Psychonomics Object Perception and Memory meeting, Los Angeles, CA.

Wiggins, D. (1980). *Sameness and substance*. Oxford: Basic Blackwell.

Wiggins, D. (1997). Sortal concepts: a reply to Xu. *Mind & Language*, *12*, 413–421.

Wright, R., & Ward, L. (1998). The control of visual attention. In R. Wright (Ed.), *Visual Attention* (pp. 132–186). New York: Oxford: University Press.

Wynn, K. (1998). Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences*, *2*, 296–303.

Xu, F. (1997). From Lot's wife to a pillar of salt: evidence that physical object is a sortal concept. *Mind & Language*, *12*, 365–392.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, *30*, 111–153.

Xu, Y. (2001). Integrating color and shape in visual short-term memory for objects with parts. Manuscript submitted for publication.

Xu, Y. (2001). Limitations in object-based feature encoding in visual short-term memory. Manuscript submitted for publication.

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340.

Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science*, *6*, 182–186.

Yantis, S., & Hillstrom, A. (1994). Stimulus-driven attentional capture: evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 95–107.