Objects, Pilots, and the Act of Attending: A Conative Account of Visual Attention[1]

by
Jerzy P. Jarmasz, B. A. Sc., M. A. Sc.

A thesis submitted for examination
to the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Department of Cognitive Science

Carleton University
Ottawa, Ontario
September 18, 2003

---

[1] Carleton University Cognitive Science Technical Report 2003-04
URL http://www.carleton.ca/iis/TechReports

The undersigned recommend to the
Faculty of Graduate Studies and Research
acceptance of the thesis

"Objects, Pilots, and the Act of Attending:
A Conative Account of Visual Attention"

Submitted by JERZY JARMASZ, M.A. Sc., B.A. Sc.,

In partial fulfillment of the requirements for the degree of Doctor of Philosophy

Chair of the Examining Board

Jo-Anne LeFevre

Institute of Interdisciplinary Studies, Cognitive Science
Department

Chris Herdman

Thesis Supervisor

External Examiner (Ph.D)

Carleton University
September 8, 2003

J. W. J.

8.IV.1939 –26.VIII.2002

Wieczny opdpoczynek racz Mu dać Panie,

a światłość wiekuista niechaj mu świeci na wieki wieków

ABSTRACT

Current research on visual attention is dominated by the object-based thesis, whereby visual input is organized into Gestalt groupings, and attention operates over these groupings. On this view, attention is more effectively allocated to a single object than to two or more. This theory is supported by experimental data, and has some ecological validity from research on Heads-Up Displays (HUDs). This research shows that under certain conditions pilots' attention becomes fixated on one object (the HUD) at the expense of monitoring events from another object (the outside scene). This phenomenon, called "cognitive tunnelling," cannot be explained by the spatial parameters of attention, and thus is consistent with the object-based thesis.

Despite the evidence for object-based attention, the conceptual foundations of the theory are questionable. Perceptual organization is an inference-based process, rather than one based on Gestalt principles. Also, evidence shows that conative factors – motivations, goals, needs – play important roles in what people perceive and attend to as objects. In the present research, a series of seven experiments was carried out to examine attentional mechanisms underlying HUD use. These experiments show that attentional strategies, task demands and the locus of control of the display strongly influence on what pilots focus attention, and to what degree. On the basis of these results, and of a critical review of the research literature on attention, a conative model of attention is proposed. On this model, attention is an active process wherein the visual system uses visual objects as tools for directing attention according to an observer's background knowledge, intentions, and task context. Observers focus on what is relevant to their tasks and needs, subject to perceptual constraints. According to this model, task demands and pilot training determine to what degree pilots "tunnel" their attention onto a HUD or its sub-parts. Wider implications for human factors research on HUD use, and for the study of cognition in general, are discussed.

ACKNOWLEDGEMENTS

There are too many people to thank. Here are but a few:

- My supervisor, Chris Herdman, for all of these years of support and encouragement, and for setting up a great lab with great facilities
- Dr. Andy Brook, Head of the Ph.D. program in Cognitive Science and last-minute committee member, whose thorough comments vastly improved (and lengthened!) my dissertation, and without whom there would have been no cognitive science at Carleton
- My external examiner, Dr. Robert McCann, who after putting up with the Great Blackout of ’03 was willing to risk a second trip to Ottawa
- The other members of my thesis committee (Drs. Vellino and Langlois) who put up with all 276 pages of my (pre-revisions) thesis (I know they did, because they found errors even in my References!)
- Dr. Lew Stelmach of the CRC, who helped me get my experiments going and who should have been on my committee
- Dr. Robert West, who once was on my committee and who gave me encouragement and advice at a crucial stage of the dissertation process
- Colleen Fulton and Lianne Dubreuil, without whose kindness, patience, cunning and teatime snacks I would have not survived the Ph.D.
- Kamilla Johannsdottir, for all of the discussions we have had about visual attention

and aviation psychology over the years, and for being there when it counted (many, many times…)

- All of the people who are or who have been members of the CACR and ACE Lab over the years; you made my everyday work and my experiments that much easier and more fun
- All of the students I have known over the years in the Ph.D. program in Cognitive Science; you forced me to think my way through a wide variety of interesting topics (only some of them cognitive), and you made sure I was having fun while doing so (except for Neville; you're just a miserable jerk)
- All the staff at Mike's Place (especially Andrew Prime and Aalya Ahmad), who provided a (relatively) safe environment for our cog sci debates and discussions
- Jacky Tweedie, who helped me sharpen many of my thoughts in the later stages of the dissertation, who reminded me about this cool dude called Vygotsky, and who for a few, too brief months reminded me to live
- My family (Mom, Mario & Lidia), who gave me the support and material environment I needed in the very last stages of writing my dissertation
- My cat Dougall, who has been keeping me sane since 1999, while evil forces were conspiring to take my sanity away

Finally, I am grateful to NSERC and CRESTech for allowing me to eat and pay rent for most of the duration of this Ph.D.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Research on visual attention is beset with confusion. This is not a new claim, to be sure, as it has been made in many ways by Spearman (1937, quoted in Wright & Ward, 1998), Allport (1989), and Fernandez-Duque and Johnson (1999), to name but a few. However, there is a particular way in which research on attention is confused which has not yet been clearly drawn out. Attention has been variously conceptualized as a filter (Broadbent, 1958), a resource (Kahneman, 1973), a spotlight (Posner, 1980), an epiphenomenon of eye movements (Sheliga, Riggio, & Rizzolati, 1995), and so on. The one common thread among these views is that attention is conceptualized as an enabling precondition for cognitively-mediated activity, as mechanism or as resource. Few if any of the accounts that have currency in experimental cognitive psychology exploit the common-sense notion of *paying attention.* Current cognitively-oriented research on attention fails to explicitly note that attention is an activity as much as a precondition for action itself. Yet the notion of paying attention creeps in through the back door with the notions of *attentional set* and *top-down attention,* meant to capture those aspects of attention – still construed as an enabler of cognition – that cannot be accounted for by stimulus properties alone.

I argue that attention is best understood as an activity that is cognitively mediated, and not a mere reflexive response to putative salient properties of visual stimuli. It has been recognized that attention seems to have a representation-mediated aspect, in the form of what has been called object-based attention (Baylis & Driver, 1998, for a review). However, the object-oriented view still conceives of attention as a reflexive response to salient features of visual stimuli that enables object recognition, mediated by background knowledge, to occur.

Taking attention to be merely a mechanism, passive (as in the filter model) or reflexive (as in the spotlight and object-oriented models), or a resource (Kahneman, 1973) for enabling action, however well it might account for experimental data generated in the laboratory, comes up short when we try to make sense of how we, as cognitive *agents*, actually deploy attention in the world, i.e., *pay attention* to things. Furthermore, how and what we pay attention to have significant implications not just for what we see (as most theories of attention allow), but also for how we see what we see, how we *construe the world, make sense of it, and act upon it.*

I will argue that studying action and cognition “in the wild” (Hutchins, 1995), requires that attention be construed as an activity, as well as an enabler of activity. In doing so I will focus on one specific domain of activity, the use of Head-Up Displays (HUDs) for flying aircraft. Studying how attention mediates the use of a HUD reveals that attention is an activity that itself shapes further activity, rather than a merely reflexive selection of visual stimuli directed by object representations. Examining how pilots use HUDs is a particularly useful domain of activity for studying attention because there are few other domains that require people to process so much information in such a time-critical manner with such high stakes.

The goal of the present research is to give a new positive account of visual attention based on the notion of attention as activity. The central claim of this account is that, in order to do what attention is generally said to do (filter information, enhance processing, integrate visual features into unified percepts), the visual system uses basic visual features (colour, shape, location, surfaces, motion, depth information, and so on) as tools in order to direct attention according to an observer's background knowledge, goals, intentions, and particular task demands. The deployment of attention itself then organizes and transforms these basic visual stimuli into a meaningful parsing of one's visual environment, through the effects noted above. In arguing for this positive account, I will start with a negative account, by discussing what attention is not.

Attention is not a reflexive mechanism "based" on any particular property or set of properties of visual stimuli. Instead of being "space-based" or "object-based," attention is space- and object-*mediated*. Spatial and object features are tools used by the visual system to deploy attention. To make this point, I will start with a review of current research on visual attention. The conception of attention as object-based – the theory that attention is allocated to perceptual units provided by various perceptual grouping mechanisms – seems to have become the dominant view in attention research, especially in applied research, and I will therefore use this view of attention as a foil to my positive account. However, the earlier view that attention is deployed primarily to regions of visual space regardless of their content – the "spatial" or "spotlight" view – still holds currency, though it is often seen as a contributor to object-based attention (Vecera, 2000), and will therefore be briefly discussed as well (Chapter 2). After this general review, I will critically review certain key assumptions of the object-based view, and I will in particular show how the view becomes problematic when it is applied to real-life stimuli, specifically to HUDs, as discussed above (Chapter 3). This critique lays the foundations for the experimental studies I carried out and describe in Chapter 4. On the basis of the critique in Chapter 3 and my experimental results in Chapter 4, I lay out my positive account of attention (Chapter 5). I discuss the implications of this account for researching attention in real-life settings, again focusing on the case of HUD use (Chapter 6). Finally, I draw out the broader implications that my view of attention has for the study of cognition in general (Chapter 7).

2. REVIEW OF RESEARCH ON VISUAL ATTENTION

2.1 Some Basic Concepts in Attention

Visual attention is a fundamental topic in cognitive psychology and a topic that has been raised extensively in human factors research. However, despite the numerous models of attention that have been proposed over the years, there is still much confusion as to the nature and role of attention. Clearly, a primary role for attention is to selectively enhance the processing of "privileged" information. But this is by no means the only role. It has been proposed that attention serves to enhance neural information processing (Stelmach, Campsall & Herdman, 1997), to modulate motor responses to stimuli (Tipper & Weaver, 1998), to maintain working memory (Baddeley, 2000; Engle, Kane, & Tuholski, 1999), and to sequence cognitive operations (Fernandez-Duque & Johnson, 1999). It is also thought that attention is necessary for the binding of perceptual features into a single phenomenal object (Treisman, 1998; Treisman et al., 1983; Treisman & Gelade, 1980). In addition, attention allows an organism to select a relevant mental representation of the environment to guide further action (Tipper & Weaver, 1998). Further away from perception, attention has also been construed as a mechanism for generating concepts by abstracting specific features from, and providing relative emphasis in, internal, perceptually-generated representations (Barsalou, 1999; Langacker, 1986).

The contrast between the Treisman (1998; Treisman et al., 1983; Treisman & Gelade, 1980) account and the Tipper and Weaver (1998) view is informative. While both assume that attention operates on mental representations, Treisman contends that attention is necessary to form coherent representations, while Tipper and Weaver assume that attention selects a given representation, and take some form of completed representation for granted. This illustrates one of the fundamental tensions in the attention literature: does attention play a role in integrating perceptual information into coherent representations, or does it simply select "pre-formed" representations for further processing? This has been expressed in the literature in many ways, the most prominent being the debates on early vs. late selection and on spatial vs. object based attention. In both of these debates, one camp assumes that attention is a pre-requisite for all perceptual processing; this is the case for the early selection and the spatial models of attention. The other camp assumes that some degree of perceptual processing occurs pre-attentively, and is indeed a pre-requisite for attentional selection. On this view, attention simply serves the purpose of selecting one representation provided by this "pre-attentive" processing; late selection and object-based attention models tend to fall into this category.

Experimental research lends some degree of support both to the notion that attention precedes all perception, and to the view that some processing must precede attention. In the following, this tension will be examined in terms of the "spotlight," and object-based accounts of visual attention. While both views are useful in understanding the processing of visual information, object-based attention theories have been invoked more successfully in the study of dynamic (i.e., more ecologically valid) visual environments, both in the laboratory (Pylyshyn & Storm, 1988) and in applied settings (Wickens & Long, 1995). Thus, the emphasis in this review will be on object-based attention.

2.2 Early Models of Attention: The Filter and the Spotlight

The earliest mechanistic model of attention is the filter model (Broadbent, 1958). This model grew mainly out of the application of information theory, as developed by Shannon (1938), to research on auditory attention. On this model, attention is a passive cognitive structure that protects the system from information overload by filtering out information. Some information is let through the filter for further processing, whereas everything else is discarded. What is filtered out is determined on the basis of the physical characteristics of the stimulus, before any kind of conceptual processing takes place (i.e., before a stimulus is categorized or identified).

While the filter model has been useful in accounting for certain aspects of auditory information processing, this model has met with less success when applied to vision. For instance, it has been shown that using a visual cue to draw a participant's attention to a particular region of a display enhances processing of information at that location (Posner, 1980). In such cases, the stimuli to be processed are often the only ones in the display, and there is therefore nothing else to filter out. Nevertheless, responses to cued stimuli are more rapid than to uncued ones. These data cannot be accounted for with a model that simply assumes a passive filtering out of irrelevant data. Rather, they suggest the visual system has the ability to selectively enhance processing in particular regions of the visual field, much in the way a spotlight illuminates a particular region of a darkened stage. This conception of visual attention came to be known as the spotlight model (Fernandez-Duque & Johnson, 1999).

2.2.1 Characteristics of the Spotlight Model

There are two main versions of the spotlight model, both of which are based on the notion that attention "highlights" a region of the visual field. One version of this model is similar to the filter model in that what falls outside of the attentional spotlight is assumed not to be processed (Posner, 1980). In the second version, the spotlight serves to concentrate attentional resources to a particular region in space, thereby enhancing processing at that location, but without completely eliminating processing of the unattended regions (Downing & Pinker, 1985; Jonides, 1981).

The spotlight models, however, differ from the filter model in several important respects. First, the attentional filter is viewed as a structure through which information must flow. In contrast, the attentional spotlight is not a structure but rather a functional enhancement of information flow. On this view, the spotlight does not passively block the processing of extraneous information. Rather, it actively directs processing resources to the area within the spotlight. Second, the spotlight has a spatial dimension that the filter lacks: the spotlight selects a region of space, whereas the filter is insensitive to the spatial layout of information. On this model, the latency to respond to two stimuli increases as the spatial separation between the stimuli increases, because the spotlight must travel from one to the other (Posner, 1980). These so-called "spatial effects" are often considered the marker effect of the spotlight model. Third, whereas the parameters of the attentional filter are assumed to be fixed (Fernandez-Duque & Johnson, 1999), the spotlight can be controlled. In particular, the spotlight can be moved to different parts of the visual field and, according to some specific models, the size and possibly the shape of the spotlight can be changed. This is particularly important as most spotlight models incorporate the assumption that the spotlight *cannot* be split between separate regions of space. Accordingly, attending to several regions of space required serial shifts of the spotlight.[2] Interestingly, problems of attentional control were not an issue with the filter model, which lacked a spatial dimension.

In sum, the spotlight models improve on the filter model by introducing the notion of enhancement of information processing (crucial to cueing data), and mechanisms for the control of attention. Note, however, that the spotlight account still construes attention as a reflexive mechanism, which functions largely by responding to salient cues in the visual field, where salience is defined in terms of stimulus properties alone (Posner, 1980).

2.3 Evidence for Object-Based Attention

The object-based attention hypothesis is based on the assumptions that perceptual (but pre-conceptual) organization of a visual scene into discrete units (generally called

---

[2] Later versions of the spotlight model include the assumption that the attentional spotlight can be split among two or three regions of space (Bichot, Cave, & Pashler, 1999; Castiello & Umiltà, 1992). However, this modification does not substantially change the spotlight account of attention.

"objects") occurs before attention is allocated, and that attention then selects or enhances visual stimuli as organized into objects rather than undifferentiated regions of visual space. On this view, it is more difficult to divide attention between two or more objects at once than to attend to a single object, even if they occupy the same region of space (i.e., the two objects overlap). This is known as the "object effect" or the "object advantage" in the research literature (see Lavie & Driver, 1996). The contrast with the spotlight model is clear. Since, on the spotlight model, everything in the spotlight is assumed to be processed in parallel, features from two nearby or overlapping objects should be attended as easily as a single object, whereas on the object-based model this would not be the case. As discussed below, there is a well-established body of evidence in support of the idea that dividing attention between objects results in less efficient processing than attending to a single object. It should be noted that spotlight and object-based attention theories are not contradictory but rather complementary (Lavie & Driver, 1996; Logan, 1996), despite the fact that they have often been described as rivals (Driver & Baylis, 1998). Nevertheless, the object-based theory accounts for many phenomena better than the spotlight model does.

2.3.1 Behavioural Evidence for Object-based Attention

Experiments by Duncan (1984), Treisman et al. (1983), and Baylis and Driver (1992, 1993; Driver & Baylis, 1989) first provided credible evidence that, under certain circumstances, attention is better described as selecting objects rather than regions of space.

Duncan (1984) presented participants with two overlapping figures, a box and a line drawn through it diagonally (Figure 1a). The two objects each could vary with regard to two properties: the box could vary in size (small or large), and in the location of a gap in its border (right or left edge); the line was either dotted or dashed (texture) or tilted to the left or to the right (orientation). The participants' task was to identify two attributes at once; in one condition they were to report on attributes from a single object (either the line or the box); in the other, they were to report on one attribute from each object. On the assumption that attention is analogous to a spotlight, it would be expected that participants' performance would not be affected by whether the target attributes were spread across one object or two, as they overlap in space. However, the results showed that participants' identification was more accurate when the two attributes were located on a single object relative to the case where each attribute was located on the other object.[3]

Treisman et al. (1983) obtained a similar cost in performance when the targets to be identified belonged to different objects. Participants were presented with a rectangular frame and a word, configured in one of two ways (Figure 1b). In one configuration, the frame and the word were presented apart (above and below a fixation point), presumably representing two distinct objects. In the other, the word was presented within the frame,

---

[3] Stimuli were followed by a mask of irregular lines to ensure that performance reflected attentional selection of visual stimuli rather than the processing of information in visual iconic memory (i.e., a mental "after-image" of the stimulus; Neisser, 1967).

assumed to form a single object.[4] In both cases the distance between the outline of the frame and the word was 1° of visual angle. The participants' task was to read the word and to simultaneously assess the location of a gap in the frame. The distance between the gap and the word was always the same, regardless of their relative locations. The results clearly showed that performance was significantly facilitated when the word was presented within the frame (presumably forming a single perceptual object) compared to when the word and the frame were separate.

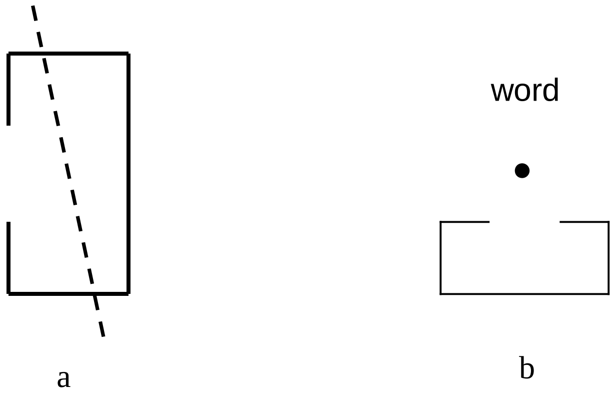

Figure 1: Examples of displays used in early research on object-based attention. (a): Duncan (1984): (b): Treisman et al. (1983)

Driver and Baylis (1989) further demonstrated the limitations of the spotlight model by using the flanker task (Eriksen & Eriksen, 1974), an experimental task that had

---

[4] The difficulties with determining how many objects are "available" for visual selection based on the number of objects assumed to appear in a display is discussed in Chapter 3.

generally been used in support of the spotlight model. The flanker task requires participants to identify or classify a target element (typically a letter) in the presence of distractors. Interference from distractors had been shown to increase as a function of the spatial proximity of the distractors to the target (Eriksen & Eriksen, 1974). Driver and Baylis used the Eriksen and Eriksen paradigm (participants responded to a central letter located in an array of five letters) but grouped the target and distractors together using common motion. When the outer letters moved with the target they interfered more with target identification than the nearby letters that had remained stationary. This demonstrated how a seemingly spatial effect breaks down when the target and the distractors are grouped by non-spatial factors. Using the perceptual grouping principle of common fate (Koffka, 1935; see also discussion below), the distant distractors were grouped with the target and as such produced more interference than distractors located closer to the target. The spatial models cannot account for these results, as the basic claim of the space-based hypothesis is that attention is allocated to contiguous regions in space, and that everything within such a region gets processed.[5]

On the strength of early results such as those described above, it was realized that the spotlight model of visual attention was inadequate for a full account of visual attention, and the object-based theory of attention was formulated (Duncan, 1984, provided one of the first statements of this thesis). As noted above, the core of these

---

[5] These results are compatible in principle with multiple-spotlight models. However, the notion of multiple spotlights is inconsistent with the notion of a unitary attentional spotlight that spatial models of attention are built on (see discussion in Fernandez-Duque & Johnson, 1999).

models is that attention operates on visual information that has already undergone some degree of processing which has organized the retinal image into objects or perceptual groups. Further experimental evidence, reviewed below, has been adduced over the years to support and develop this hypothesis. Before reviewing these data, it is worth explicating the notion of object generally subscribed to in attention research.

It has been generally assumed (rather uncritically) that the objects that attention selects are perceptual groupings generated by (or that are consistent with) the principles of perceptual organization as proposed and studied by the school of Gestalt psychology (see Driver & Baylis, 1998 for an explicit statement of this position). While these principles are used extensively in the attentional literature, current opinion in research on perception is critical of Gestalt theory in general as an explanatory model for perceptual organization (Hoffman, 1998; Palmer, 1999). Nevertheless, the following grouping principles are commonly used in research on object-based attention: grouping by similarity (elements of a display sharing a common attribute, such as shape or colour, are generally perceived as being grouped together); grouping by proximity (elements which are close to one another are generally perceived as forming a group); grouping by continuity, completion and closure (elements which only partly suggest a complete shape, such as collinear dashes, the corners of a box, or part of a disk, are generally perceived as forming the completed shape); and grouping by common fate (elements which move together are perceived as forming a group). Treisman et al.’s (1983) and Duncan’s (1984) displays seem to rely mainly on the principles of completion and closure, whereas the stimuli used by Driver and Baylis (1989) rely on the principle of common fate.

As noted, many studies have developed the theory that attention is object-based, showing the many ways in which various attention-sensitive tasks are dependent on (Gestalt-type) perceptual organization. Goldsmith (1998) showed that visual search is easier when features are linked to the same object than when they belong to different objects. Similarly, Duncan and Nimmo-Smith (1996) found that it is harder for participants to discriminate between features that belong to different objects compared to a situation requiring discrimination between features that belong to a single object. Kramer and Jacobson (1991) used a variation of the response competition paradigm (Eriksen & Eriksen, 1974) to test the object effect in a focused attention task. Participants judged whether the target (a line) was dashed or dotted while ignoring distractors. Distracting lines (either compatible or incompatible with the target) were located to the left or right of the target. The distractors could be grouped according to the Gestalt principles with the target, or they could form a part of a different object. The distance between the target and the distractors was kept constant. According to the space-based models there should be no difference between the different conditions, because the spatial separation was constant. On the other hand, according to the object-based model there should be less interference when the distractors belong to a separate object. As predicted by the object-based model, the interference from distractors was drastically reduced or eliminated when the distractors and the target belonged to different objects. Further, when target and distractors were grouped into one object, there was clear evidence of interference, which varied as a function of the compatibility between distractor and target (incompatible distractors produced significantly worse performance relative to compatible ones). Note that the interference from distractors seen in these studies cannot be explained by referring to the spatial hypothesis. Grouping the distractors and the target together by colour or good continuation causes significantly more interference with target response than distractors that are easily separated from the target. Spatial distance seems to have no influence here. This suggests that visual attention is directed to perceptual objects in the visual field that are segmented according to the Gestalt grouping principles.

Experimental evidence supporting object-based attention has also been adduced from inhibition-of-return (IOR) tasks. IOR was originally interpreted as supporting spatial models of attention, as certain locations in space are prevented from being constantly re-examined (Posner & Cohen, 1984; see also Tipper & Weaver, 1998, for a review). However, if attention is object-based then the inhibitory mechanism should be referenced to perceptual groups regardless of their position, rather than to specific locations in visual space. To this end, a number of studies suggest that IOR is related to perceptual objects in the visual field rather than spatial location (Tipper, Driver & Weaver, 1991; Tipper & Weaver, 1998). Jordan and Tipper (1998) used a static display to examine the difference between cueing locations vs. cueing visible objects in the display. The display consisted of black “pacmen” (discs with a quadrant missing) and lines. In one condition the “pacmen” were oriented so as to form illusory squares (so-called Kanizsa squares; see Figure 2), whereas in another condition the “pacmen”, while occupying the same location as in the first condition, were not oriented so as to form illusory contours. Thus, in the first condition cueing a particular location would coincide with cueing an object, whereas in the second condition, cueing the same location would not cue an object, while maintaining the other attributes of the display the same across conditions. The IOR effect was much stronger in the case where a spatial cue also cued an object, thus suggesting that some portion of the IOR effect is referenced to objects and not just their location.

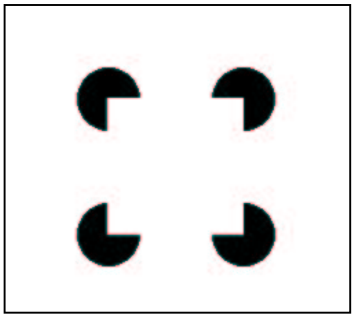

Figure 2: A Kanizsa square as used by Tipper and Weaver (1998)

Perhaps the most compelling evidence for object-based attention comes from research done with moving displays. As we have seen, Driver and Baylis (1989) had already shown that common fate is a powerful grouping principle affecting attention. Along similar lines, Tipper and Weaver (1998) carried out IOR experiments using elements endowed with common motion. Research on IOR with static displays easily lends itself to a spatial explanation: attending to a particular location can inhibit attending to the same location a little while later. However, Tipper and Weaver reasoned that if an organism is to successfully survive in a dynamic and complex environment, then the IOR phenomenon must also apply to moving objects. In this case, it wouldn't do to inhibit return to a specific location, because, by definition, moving objects are changing their location. Thus, it would make sense to expect that an object-based IOR mechanism exists to allow efficient scanning of scenes involving moving objects. Tipper and Weaver's data do in fact suggest that there is an object-based component to IOR, in that participants were slower to return attention to a pre-cued moving object. While their data do not exclude the possibility that IOR is at least in part location-based as well, they certainly argue for the notion that for attention to be allocated to moving stimuli, it must also be object-based.

Also using displays of moving elements, Pylyshyn and Storm (1988) tested the assumption, arising from a spatial model of attention, that people are able to track many elements by moving their attentional spotlight from element to element in rapid succession. They first tested participants' ability to track a subset of elements within a display of elements all endowed with independent trajectories. This showed that people are able to successfully track about four or five elements with independent motion for at least 10 seconds. A computer simulation of a single attentional spotlight tracking multiple objects with the same trajectories as the ones shown to participants was developed. This simulation showed that an attentional spotlight moving from element to element would not be able to keep track of the elements as efficiently as did the actual participants. Pylyshyn and Storm concluded that a spatial attention hypothesis cannot account for people's ability to perform multiple object tracking.

Two competing object-based accounts have been proposed of how people might perform multiple tracking of objects. One account is that the multiple elements are formed into a nonrigid polygon, with each element being one of the vertices of the polygon. On this view, Yantis (1992) found that tracking performance is affected by factors that facilitate the initial formation and maintenance of a perceptual group of elements to be tracked. Thus, Yantis argues that specific elements within a display of similar moving elements are tracked by grouping the elements into a single “superobject” (essentially a nonrigid polygon). The less the nonrigid polygon resembles a regular, rigid polygon, the harder the task.

Pylyshyn and his colleagues have developed a different account, which assumes that each element being tracked is associated to a visual index, or a FInger of INSTantiation (FINST; Pylyshyn, 1989). On this view, the early visual system attaches indexes to the individual elements to be tracked. These indexes provide a way for the visual system to pick out specific elements of the visual field by referring to the elements themselves, and not to any properties of the objects (Pylyshyn, 1998). These indexes then allow the rest of the visual system to attend to those specific elements, to track them, to identify them, and so on (Scholl & Pylyshyn, 1999; Sears & Pylyshyn, 2000). Thus, visual indexes are a kind of representation or data structure within the visual system which functions in a manner analogous to linguistic indexes and demonstratives (e.g., words such as “that” or “there”).

In sum, the ability to track moving objects appears to require that attention be object-based at least to some degree. Spatial models of attention are inadequate for explaining the experimental evidence obtained from multiple tracking tasks, as these models would require the spatial spotlight to visit each moving element in rapid succession. Successive tracking of this sort would mean that the visual system is able to predict the positions of the moving elements and to move the attentional spotlight at a speed that is beyond the capacities of the human visual system.

2.3.2 Evidence from Aviation Psychology

Research in aviation psychology, in particular research bearing on the use of HUDs, is a domain of applied research that is particularly well suited to the study of visual attention in a real-world setting. This is because pilots are increasingly expected to process large amounts of visual information, much of it generated within the cockpit by computers. HUDs are a case in point, as they provide pilots with computer-generated graphical representations of information about the outside environment and the aircraft's status. It is therefore not surprising that researchers investigating HUD use have seized on object-based attention as an explanatory construct. As discussed below, this applied research has shown that object-based attention is a construct that has a measure of external and ecological validity. Furthermore, object-based attention has been used to suggest ways in which HUDs can be improved. The aviation psychology research literature thus provides evidence that object-based attention is a very useful and powerful theory. Thus, as a review of relevant literature will show, research on object-based attention has benefited from being used in the applied domain of aviation psychology, and aviation psychology has itself benefited from the theory of object-based attention.

In an early study of attentional effects in HUDs, Fischer et al. (1980) described a simulator-based experiment where pilots were required to perform runway approaches flying an aircraft that was equipped with a HUD versus a traditional Heads Down Display

(HDD). It was found that some of the pilots using the HUD failed to notice unexpected intrusions on the runway when they were also required to attend to events in the near domain (see also McCann & Foyle, 1995). This failure to notice runway intrusions was not experienced by pilots using the HDD. Although the results suggest that the presence of the HUD is responsible for the degraded performance, the Fischer et al. study was flawed in that the location of the instrumentation (HUD vs. HDD) was confounded with the type of instrumentation.

Wickens and Long (1995) repeated the Fischer et al. (1980) experiment but with matched instrumentation across the HUD and HDD. In contrast to the Fisher et al. findings, pilots using the HUD were successful in noticing runway intrusions. However, these pilots were considerably (2.5 seconds) slower to respond to intrusions than were pilots using the HDD. In sum, both the Fisher et al. (detection accuracy) and the Wickens and Long (detection time) studies show a disadvantage for HUDs versus HDDs under certain conditions. This disadvantage seems to arise in situations where, prior to the unexpected intrusion, the HUD is the element of the display most relevant to the task.

The disadvantages of HUDs are further illustrated in a study by Foyle, Sanford and McCann (1991), which required pilots to control their flight-path while maintaining a fixed altitude. Superimposing a HUD digital readout of altitude onto the flight path resulted in excellent control of altitude. However, when focusing on altitude, pilots tended to collide with the flight-path markers, such as buildings or landmarks. This tradeoff between using the HUD symbology (digital altitude) and processing of the external scene cannot be attributed to pilots' needing to re-accommodate visual focus when switching between visual domains: both the HUD and the outside scene were presented at the same optical distance. Instead, this evidence suggests that when the HUD symbology is required for performance, the symbology attracts the pilot's attention at the cost of attending less to objects and events in the environment. That is, when focusing attention on one domain, information located on the other domain seems to be processed less effectively. This phenomenon has been labelled cognitive tunnelling (Martin-Emerson & Wickens, 1997; Wickens & Long, 1995), or, more recently, attentional tunnelling[6] (Fadden, Ververs & Wickens, 2001).

The evidence for object-based attention has led researchers to claim that the cognitive tunnelling experienced with HUDs is due to the near and far domains forming separate perceptual groups or objects (see Foyle et al. 1991, and McCann, Foyle & Johnston, 1993, for explicit statements to this effect). This claim is predicated on the notion that near and far domains differ along one or more of the Gestalt grouping principles (Foyle et al., 1991; McCann et al., 1993). In particular, the HUD symbology is stationary relative to the pilot- or aircraft-centric view, whereas the external scene is in constant motion (grouping by common fate). Also, HUD symbology is usually displayed in a uniform colour, which may differ from the various colours of the external scene (grouping by colour). The relative size and self-containedness of the HUD might also

---

[6] It is interesting to note the switch in terminology. One can only guess at the reasons, as none was given in the Fadden, Ververs and Wickens paper, but I would suggest the authors felt the term "cognitive tunnelling" implies too large a role for conceptual, "top-down" knowledge in attentional control.

lead pilots to perceive the HUD as figure and the outside scene as ground (figure/ground organization).

The claim that the near (HUD) and far (external scene) domains form separate perceptual groups provides a possible basis for cognitive tunnelling when combined with the object-based hypothesis. On this view, when pilots attend to the near domain, all of the HUD symbols get processed quickly in parallel while processing of information in the far domain is delayed. Assuming that attention can be asymmetrically distributed between objects so as to be focused almost exclusively on one only, cognitive tunnelling could be the result of attention being allocated in an object-based manner.[7]

In addition to providing possible explanations for attentional difficulties in using HUDs, object-based models of visual attention have also been used to suggest modification in HUD design in order to alleviate these problems. The proposed solutions to cognitive tunnelling have generally aimed at improving pilots' ability to integrate information from the near and far domains by fusing both domains (or at least some aspects of both) into a single domain or perceptual group. One way of doing this is to use conformal symbology. Broadly speaking, the definition of conformality used by HUD developers refers to the degree to which a symbol forms an object within the scenery. The idea is that a conformal symbol should serve as a virtual analog for far domain elements. In other words, symbology that is an accurate graphic representation of an actual object represented in the far domain, or that forms a one-to-one correspondence with the world

---

[7] Experimental evidence for asymmetrical distribution of object-based attention is provided in Chapter 4.

is deemed to be conformal (Martin-Emerson & Wickens, 1997). On this view, conformal symbology can be a virtual runway overlaying the actual runway or a scene-linked symbology element where, for example, altitude is represented within an object inserted into the far domain. Non-conformal symbology would be symbols such as a digital readout of the altitude or airspeed, path guidance information like glide slope, or localizer symbology (Wickens & Long, 1995). It should be noted that according to this definition, even traditional HUDs have some conformal symbology (e.g., a horizon line). In contrast, symbols representing vertical speed, airspeed, distance, and altitude etc. are usually non-conformal. A similar approach consists in using scene-linked symbology (Shelden, Foyle & McCann, 1997), whereby flight information (e.g., airspeed) is represented on virtual objects added to the outside scene.

Experiments examining conformal and scene-linked symbology in HUDs have yielded promising results. For example, research by Foyle, Sanford and McCann (1991; see also McCann & Foyle, 1994, and Shelden, Foyle & McCann, 1997) has shown that when using scene-linked symbology pilots are able to maintain altitude and follow a flight path without significant trade-offs in performance. In these studies, altitude symbology was fused with the outside scene by placing the symbology on virtual buildings along the flight path. In contrast, when the altitude indicator was merely superimposed onto the outside scene (thus, non-conformal), the task of maintaining altitude reduced flight path performance. Similarly, a study by Wickens and Long (1995) demonstrated that a HUD with conformal landing symbology (a runway overlay) was subject to less cognitive tunnelling than a HUD with non-conformal symbology, in that pilots had better awareness of runway intrusions with the conformal symbology.

Varying the manner in which conformal symbology represents information does not seem to diminish the enhanced performance. McCann and Foyle (1995) and Shelden, Foyle and McCann (1997) have shown evidence for the same benefit of conformal symbology over non-conformal symbology regardless of whether the form of the symbology was analog ("clockface") or digital. These experiments are promising and suggest that the conformal character of the HUD symbology enables parallel processing of information from the two domains. In accord with an object-based hypothesis, conformal symbology might allow for the creation of a single far domain (or object layer) of information. On this view, performance is enhanced because the pilot is able to allocate attention to the far domain without the need to switch to the near domain.

In sum, an object-based approach can be useful in assessing the attentional problems associated with the use of HUDs (i.e., cognitive tunnelling). It can also suggest ways of dealing with these problems (i.e., conformal and scene-linked symbology). Thus, the study of attentional issues in HUDs provides a degree of ecological validity for the object-based account of visual attention.

2.4 Interim Summary I: Does the Object-Based Theory Work?

The studies reviewed above conclusively demonstrate that, under certain conditions at least, visual attention is distributed according to the perceptual structure of visual stimuli, and not simply directed to regions of space. Furthermore, the hypothesis that attention is allocated to visual objects also seems to have explanatory power in HUD research, in particular regarding cognitive tunnelling, and to provide a useful design heuristic in the form of conformal symbology. What is less clear is whether it is fair to call attention object-*based*. That is, does the available evidence warrant the claim that the perceptual organization of a scene into objects determines and directs the distribution of visual attention? Or is it rather the case that many factors determine how attention is distributed, and visual objects are but one type of factor that is involved in the whole process? There is no clear answer to these questions on the evidence reviewed above, as the experimental research on attention has involved mainly the manipulation of spatial factors and Gestalt-type perceptual grouping factors.

There are also some more immediate methodological concerns with the hypothesis that attention is object-based. Clearly, such a claim requires an account of what is meant by ‘object’. As already noted, most studies have defined objects as visual stimuli grouped together according to the Gestalt grouping principles. This approach to perceptual organization has fallen out of favour in current research on object perception (see Hoffman, 1998, and Palmer, 1999, for reviews) – is it adequate for operationalizing objects for the purposes of studying attention? Related to this concern is the fact that the limit conditions of object-based attention have not been defined. On the assumption that attention operates in a spatial medium, and perhaps in other media as well (e.g., feature space; see Treisman, 1998), in addition to the object-based frame of reference, under what conditions does attention become object-based, space-based, or based in some other medium? Or if we assume that the so-called spatial and object-based aspects are different manifestations of the same attentional mechanisms, under what conditions does attention "behave" in one manner or the other? And how would these questions be investigated – what criteria would be used to distinguish spatial and object-based attention?

In sum, visual attention does, at the very least, appear to be object-based under certain conditions; but many questions about the nature of object-based attention, the conditions under which it is manifested, and how to study these questions remain. A closer look at research on visual attention is warranted in order to determine whether the object-based theory of attention can meet these concerns, and whether current experimental paradigms can be used and extended to answer these questions.

3. A CRITICAL REVIEW OF THE OBJECT-BASED THESIS

It can be said with a fair degree of certainty that under certain experimental conditions, the allocation of attention depends on spatial properties of visual stimuli, and that under other experimental conditions, factors of perceptual organization play a more dominant role in distributing attention. We have also seen that observations in less constrained, more realistic settings, namely, research on cockpit HUDs, support the controlled laboratory findings. Following this, attention has been said to be space-based, or object-based, or some combination of both. In fact, while both theories are still sometimes presented as rivals (see discussion in Driver & Baylis, 1998), there is some agreement that they are in fact complementary accounts (Lavie & Driver, 1996; Logan, 1996), even though spatial attention is being increasingly understood as a mechanism subserving object-based attention (Vecera, 2000). At any rate, the refinement and apparent success of the object-based theory has not led to the complete abandonment of the spatial models. This raises a number of theoretical and methodological issues.

At the heart of the concerns discussed below is the problem of what exactly is meant by ‘visual object’. This question has both theoretical and methodological dimensions. The theoretical questions have to do with how the visual system is thought to organize visual stimuli into the objects used by attention: are Gestalt principles, or other low-level mechanisms that are independent of higher-level conceptual knowledge, sufficient, or are an observer’s background knowledge and current mental states also required? The methodological issues have to do with how to operationalize the construct of visual object so as to, on the one hand, allow a principled distinction between spatial and object-based factors, and on the other, not rely on the notion of object-based attention itself (that is, how to avoid the trap of defining an object as that which is selected by visual attention).

The thesis that attention is object-based relies on certain assumed answers to the questions above, though these assumptions are often tacit, and rely on the wider theoretical commitments researchers have as to the nature and function of attention and cognition in general. The following sections lay out these assumptions as they are found in the available studies on object-based attention, and consider the validity of these assumptions. Four issues are considered below: The nature of the 'objects' involved in visual attention; the relation between perceptual organization and attention; the individuation of objects in experimental displays and HUDs; and the confounding of spatial and object-based factors. This examination will show that the experimental data on attention in fact do not support the assumptions that were used in designing the experiments that generated the data.

3.1 On the Nature of Objects[8]

If attention is object-based, it must perforce select objects. But what are objects? We are surrounded by all kinds of objects. For instance, as I type these words, I see a

---

[8] Many of the remarks on objects in this research have been inspired by Smith (1996). However, as this influence is often tangential and indirect, no specific citations to Smith (1996) are given.

computer monitor, a keyboard, a mouse, my cell phone, a pen, and a glass on my desk. We are so familiar with objects it would seem we shouldn't need to define them.

Yet, when psychologists speak of object-based attention, do they mean the same things that we mean by objects in common parlance? Certainly, psychologists mean to say that visual attention selects visual stimuli that *correspond* to common objects, but it is less clear that the objects selected by attention are objects in the world. The simple geometric figures used by most research on object-based attention (see Figure 1, p. 12), and the HUDs studies by aviation psychologists make for strange "objects."

A distinction commonly made in metaphysics is between so-called concrete particulars, i.e., the individual physical objects that "really exist" in the world (Loux, 1998), and "intentional objects," which are supposed to be the objects, whether existent or non-existent, that intentional mental processes are "about" or point to (Husserl & Brentano as discussed in Hammond, Howarth and Keat, 1991). Are either of these types of objects the "objects" of object-based attention? On the assumption, widely accepted in cognitive psychology, that attention selects mental representations, concrete particulars cannot be the objects that attention is supposed to select. The issue is less clear with intentional objects.

One reading of 'intentional objects' has it that they are the objects of phenomenal, and therefore conscious, experience. On this reading, they cannot be the objects of attention, since the objects of attention are supposed to be generated by vision before phenomenal experience occurs, precisely in order for phenomenal experience of objects to happen (Mack & Rock, 1998). However, the concept of intentionality most likely had a simpler meaning, less dependent on conscious experience, for Brentano and Husserl: it most likely was meant to simply capture the "other-directed" (deictic) aspect of many mental processes (Hammond et al., 1991).[9] On this reading, the objects of visual attention could be thought of as intentional objects, in that visual attention is clearly an intentional process – it is always directed onto something – and what it is directed onto are the products of perceptual organization, Gestalt or otherwise. Nevertheless, the intentional objects of attentional selection are clearly distinct from the intentional objects of phenomenal experience, which are supposed to refer semantically to objects in the world (call this "semantically intentionality"). The objects of attention, on the other hand, while "representing" concrete objects for visual processes, do not refer to real objects in a semantic sense, they refer (or rather, "point") to them only by being in lawful relations with them (call this "simple intentionality"). Furthermore, assuming that the objects of attention and the objects of phenomenal experience are one and the same would be to conflate the notions of attention and conscious processing, a danger discussed by Hardcastle (2003) and in Section 5.2.3 below. Therefore, let us for now make a three-way distinction between c-objects (concrete particulars), a-objects (the intentional objects involved in attentional selection) and p-objects (the intentional objects of phenomenal experience). We can now note that while research on object-based attention uses "object" equivocally, referring, sometimes ambiguously, now to c-objects, now to a-objects, now

---

[9] While the original sense of "intentionality" was probably simpler, Husserl's conception of intentionality is in fact quite complex. For him, intentionality has structure; it is the filter of accumulated conscious and body knowledge through which we direct ourselves onto the world, and through which our experience of the world is shaped. But this does not mean that every intentional phenomenon or structure necessarily refers, in a semantic sense, to that which it is "about" (Hammond et al., 1991).

to p-objects, the most coherent reading of the object-based hypothesis is that attention selects a-objects as a means of constructing p-objects, through which we gain phenomenal access to c-objects.

There is, unfortunately, a further complication when it comes to tracking the meaning of “object” in vision. Much research in psychology uses pictorial depictions or virtual representations of real objects, some of which are simplified figures (refer again to Figures 1 and 2 above), while others are photographic or video reproductions of concrete objects. In every case, the two-dimensional stimulus is interpreted by the observer as depicting a real object (even if the depiction is false or depicts what would be an impossible state of affairs, as with visual illusions). No one sees a photograph of their brother and thinks “there’s a set of flat coloured surfaces that just happens to look like my brother,” but rather “there is a picture of my brother.” Thus, such images induce the phenomenal perception of concrete objects in the absence of those concrete objects, and sometimes even despite the fact that the intended concrete object does not exist. Put another way, these are 2-D stimuli that give rise to p-objects of the same type as do c-objects. Since the vast majority of research on visual attention makes use of these images, as stand-ins for real objects, it is useful to also give them a name – v-objects (for “virtual objects”).

We have a four-way taxonomy of objects that can play a role in vision: c-objects (concrete particulars), v-objects (2-D devices that are perceived as c-objects), a-objects (intentional objects selected by attention), and p-objects (the objects of phenomenal experience). The object-based attention thesis can now be restated thus: the attentional system selects a-objects in order to create p-objects, which are supposed to allow a person to know about and act upon the c- and v-objects that gave rise to the a-objects. The "objects" of object-based attention are thus a-objects, the "simply intentional" objects of the discussion above.

What, then, are a-objects? The standard answer given by workers researching object-based attention is that a-objects are perceptual groupings whose formation is governed by the Gestalt principles of perceptual organization (Kanizsa, 1979; Koffka, 1935). This definition has been explicitly supported by Baylis and Driver (1989; 1992; 1993; see also Lavie & Driver, 1996), and has been endorsed by other researchers (see Wickens & Long, 1994, 1995 among others for applications in aviation psychology). Until the advent of cognitive psychology, the Gestalt principles constituted the only available theory of perceptual organization, and were thus integrated into cognitive psychology by Neisser (1967). The choice of Gestalt groupings for a-objects, the objects that are selected by attention, was thus a natural one for object-based attention. However, as is discussed next, Gestalt groupings are an unsatisfactory account of the nature of a-objects. The discussion of what a-objects actually are is deferred until Chapter 5, where a new model for visual attention is elaborated.

3.2 Why Visual Objects Can't be Gestalt Groupings

On the Gestalt school account of perceptual organization, the human visual system tends to organize stimuli into perceptual groups according to certain principles of spatial organization. A few representative examples of these principles are: similarity;

proximity; closure; good continuation; common fate (see Figures 3 and 4). These principles are motivated by an underlying overarching principle of “figural goodness,” or *Prägnanz* as it is called by Gestalt psychologists. This principle states that the visual system organizes stimuli so as to form the simplest, most elegant, and intuitively most satisfying configuration. All the other principles are supposed to be particular instances of this one.

```
o x o x o   * *  * *  * *
o x o x o   * *  * *  * *
o x o x o   * *  * *  * *
o x o x o   * *  * *  * *
```

Figure 3: Grouping into columns by similarity and proximity

```
( )( )( )( )( )( )

( )( )( )( )( )( )

( )( )( )( )( )( )

( )( )( )( )( )( )
```

Figure 4: Grouping by closure

While these principles are useful in describing perceptual grouping phenomena, particularly in carefully contrived situations, collectively, they have three flaws which make them inadequate as an account of perceptual organization, and of a-objects. The first is that the principle of *Prägnanz*, while intuitively appealing, is too vague and subjective to support a principled account of how stimulus properties drive perceptual organization by the visual system. Thus, as Palmer (1999) states, the Gestalt grouping principles are little more than *ceteris paribus* rules: particular rules apply only in particular contexts, when they are clearly the only rule that can apply. When two or more grouping principles are applicable at once, there is no way of predicting *a priori* which principle will in fact determine the configuration the observer sees (though some work has been done studying the interaction between the various grouping principles; see Baylis & Driver, 1992). The ultimate determinant of “figural goodness” becomes the observer’s subjective experience, and whichever set of grouping principles correspond to a particular observer’s subjective experience is claimed to explain a particular perceptual configuration. In terms of the taxonomy of objects developed in Section 3.1, Gestalt perceptual organization conflates a-objects and p-objects, or conscious perception and attentional processing.

The second flaw of the Gestalt principles is that they do not provide a distinction between perceptual groups and objects, which could themselves be comprised of many perceptual groups. There is no way of telling, on the Gestalt account, whether elements within a group are parts of an object or objects in their own right, and whether a perceptual group is one object or a group of objects. The Gestalt principles provide no criteria for individuating a-objects. If a visual scene contains many Gestalt groups, themselves made up of groups, an object-oriented account of visual attention based purely on Gestalt principles could not explain to which level of grouping attention is allocated.

The third problem with the Gestalt principles is that they do not allow for a clean separation of so-called “spatial” and “object-based” effects in visual attention. It is assumed in spotlight models that multiple stimuli are inspected serially, and thus the time it takes to attend to multiple stimuli is a function of their spatial separations. Conversely, one would predict from object-based models that if multiple stimuli are located within a single object or perceptual group, they are processed in parallel, and thus the time it takes to process stimuli is independent of their spatial separations. However, the Gestalt principle of grouping by proximity predicts that the closer elements are, the more likely they are to be perceived as belonging to a single group or object. Thus, the rapid processing of elements that are close together could be explained either in terms of the spotlight model or the object-based model (Lavie & Driver, 1996).

In addition to the limitations of the Gestalt principles as an account of perceptual grouping, there are problems in the way the Gestalt principles have been applied in visual attention. Most researchers studying object-based attention make the assumptions that visual processing happens in serial stages and that the stage of perceptual organization must occur prior to the stage of attentional selection (Baylis & Driver, 1998). This is in conflict with the first problem identified above, namely, that the Gestalt principles reflect the conscious perception of grouping, which, on the serial view, can only happen after attentional selection (of Gestalt groups) occurs. Furthermore, as is discussed below, there is evidence to suggest that perceptual organization does not always precede attentional selection.

Another problem for the application of the Gestalt principles is that many researchers on vision “implicitly assume that the purpose of the visual system is to construct some sort of internal model of the world outside – a kind of simulacrum of the real thing, which can then serve as the perceptual foundation for all visually derived thought and action” (Milner & Goodale, 1998). Thus, the perceptual structures generated in the visual system are supposed to correspond to physical structures in the world; that is, there is supposed to be an isomorphism between p-objects and c-objects, and by extension, between a-objects and p-objects. As a result, the object that is supposed to be selected by attention is identified in terms of the physical properties of the stimulus presented to the visual system. Unfortunately, just as it is problematic to individuate a-objects on the Gestalt principles, individuating c-objects and v-objects is no simple matter either. Figure 5, which represents a HUD, makes this point clear: is the whole HUD a single, compound object? Is it a group of objects? Is it both? Deciding which a-objects are selected by attention on the basis of (in this case) the v-objects that are presented to an observer is a questionable enterprise.

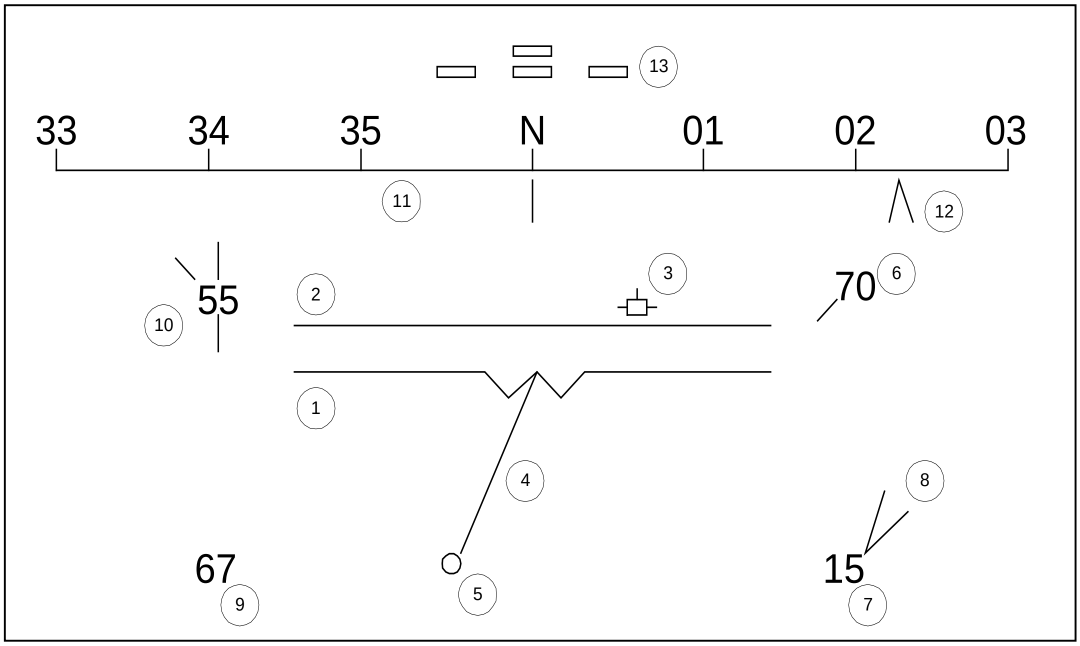

Figure 5: A HUD. Does it look like a single perceptual group?

The problem with assuming an isomorphic correspondence between the objects of visual attention and objects in the world was not salient in early experiments on object-based attention (Driver & Baylis, 1989; Duncan, 1984; Treisman, Kahneman & Burkell, 1983). The "objects" used in those experiments were quite simple, and contrived to form relatively unambiguous objects (see Figure 1 above). Thus, the objects and groups perceived in these displays were readily accounted for by the Gestalt principles of perceptual grouping. However, the need for better definitions of objects and perceptual groups is obvious when examining the role of object-based attention in more rich and dynamic contexts. For example, as discussed above, object-based attention has been used to explain the phenomenon of cognitive tunnelling. It has been suggested in many studies (e.g., Foyle, Sanford, & McCann, 1991; McCann, Foyle & Johnston, 1993; Wickens & Long, 1995) that the HUD and the outside scenery form two separate perceptual groups.

By assuming that attention is object-based, it is claimed that pilots can attend to the HUD or the scenery, but not to both at once, thus explaining cognitive tunnelling. However, as is clear from Figure 5, one cannot simply assume that a HUD constitutes a single object, much less that it is perceived as a single object based on the Gestalt principles. Thus, there is reason to question the conclusions drawn from certain experimental results on attention as to how many objects were perceived by a participant. This issue is discussed in Section 3.4.

Finally, it should be noted that, with a few exceptions (e.g., Gibson & Egeth, 1994; Lavie & Driver, 1996), most research on object-based attention fails to take measures to distinguish between spatial and (spatially invariant) object-based factors in the allocation of attention. Section 3.5 is a review of how many results purportedly supporting the object-based thesis are in fact ambiguous, and can be re-interpreted in a manner consistent with spatial attention.

3.3 Which Comes First, Perceptual Organization or Attention?

It is generally assumed in research on object-based attention that perceptual grouping is pre-attentive (see Driver & Baylis, 1998; Feldman, 1999). On this view, grouping does not require, and happens prior to the allocation of, visual attention. This has been one of the key assumptions of the object-based model. For attention to be allocated preferentially to objects, these objects must presumably be made available to the attentional system. As noted above, the objects implicated in visual attention are usually defined in terms of mechanisms of perceptual grouping. Thus, perceptual grouping is thought to be a prerequisite, logically and temporally, to attentional processing.

The view that perceptual grouping is pre-attentive has been challenged. Mack, Tang, Tuma, Kahn, & Rock (1992) report a series of studies in which participants were required to focus on a cross in the middle of a patterned display and report on the length of the arms of the cross. As a secondary task, participants were required to report on the perceptual grouping in the patterned display. The background patterns consisted of small squares organized either into rows or columns. This was achieved using grouping by proximity or grouping by colour. Mack et al. found that when participants were required to focus on the cross in the centre of the display, their ability to identify the grouping pattern, and even to detect any grouping at all, decreased significantly. The cross and the background pattern occupied the same location in the display, in that the cross clearly overlaid the background pattern. The displays had an average diameter of approximately 7.6°, while the visual angle of the squares in the pattern was typically 0.37° by 0.37°, and the spatial separation between the squares tended to be in the 0.3° to 0.9° range. Thus, it cannot be argued that the patterns were so small that participants were unable to see them, nor that they were so large that participants were unable to visually integrate the pattern due to drop-off of visual acuity. What these results suggest is that it is not enough for a grouping pattern to be foveated. If a pattern is not specifically attended, it will not be identified or even detected.

One might argue that grouping patterns are identified pre-attentively and then attentionally filtered out by focusing on a particular aspect of a display. Ben-Av et al.

(1992) have looked into whether perceptual grouping requires attentional resources, or whether these processes are inhibited by attentional mechanisms. They employed a dual-task paradigm, in which participants were required to carry out a grouping pattern identification and/or detection task concurrently with a singularity detection task. The grouping patterns were based on grouping by proximity and grouping by similarity of shape. The singularity detection task involved detecting the presence of a single predetermined shape that did not match the shapes used in the grouping patterns. One task was always designated as primary, the other as secondary. The relative performance on each task was measured as a function of masking at different stimulus-onset asynchronies (SOAs) and of which task was designated as primary. Ben-Av et al. found that at long SOAs, participants reached optimal performance on both concurrent tasks, regardless of which task was primary. However, when singularity detection was the primary task, participants' performance on the grouping tasks was significantly reduced at short SOAs. Conversely, when one of the grouping tasks was primary, performance on neither task suffered significantly at short SOAs. Thus, the results of the experiments where singularity detection was the primary task replicate the findings of Mack, Tang, Tuma, Kahn, & Rock (1992), whereas the experiments where the grouping tasks were primary suggest that focusing attention on one aspect of a display is not enough to interfere with the performance of another task. Focusing on the grouping pattern did not interfere with the task of detecting a singularity, but focusing on the singularity interfered with the detection of a grouping pattern. These findings undermine the suggestion that perceptual grouping happens pre-attentively and is simply filtered out by focusing attention on the similarity, for, by the same reasoning, the singularity should also be filtered out when the grouping pattern is focused. The more likely explanation is that perceptual grouping does in fact require visual attention, whereas singularity detection doesn't. Thus, deliberately focusing attention on the singularity starves perceptual grouping of the attentional resources it needs, whereas focusing on the grouping pattern does not interfere with singularity detection, as this task doesn't require attention.

What, if anything, is perceived without attention? Rock, Linnett, Grant, & Mack (1992) examined this question. They used a method that was similar to that used in Mack et al. (1992). The main task was again to determine which arm of a cross was longer. A number of secondary tasks were used, including the detection of a single stimulus in the background, counting the number of elements presented in the background, and identifying some attribute of a single element, such as its shape or colour. Rock et al. (1992) found that when participants were focusing on the cross (i.e., the primary task) and were exposed to the background stimulus for the first time, they were able to report the number of elements in the stimulus, as well as report their colour. However, they were unable to identify the shape of the elements that were presented. This suggests that a few basic features of visual stimuli, such as their existence, location, and colour, are processed preattentively, but that the perception of shape requires attention.

As noted above, much of the research on object-based attention relies on the assumption that perceptual grouping is a pre-requisite for visual attention. There is, however, much evidence that undermines this assumption. It should be noted that there is another influential view on object-based attention that holds that visual attention is required for perceptual wholes to be formed, not the other way around. Perhaps the best exponent of this view in recent times is Treisman (1988, 1998; Treisman & Gelade, 1980). In brief, Treisman suggests that the early visual system analyses stimuli into basic features, (e.g., shape, colour, etc.), and that attention is required to bind these features into a single percept at a specific location in space (but see Goldsmith, 1998, for an account of object-based feature integration). Here, attention is viewed as a kind of spatial map that integrates information from different feature maps by binding features that share the same location in visual space. Supporting evidence for this theory comes from visual search experiments, where a single element in a display is found much more quickly, and in a sense "pops out," when it is distinguished from all the other elements by having a unique feature (e.g., it is the only element with a given colour or shape) than when it is distinguished by a unique combination of features which are found elsewhere in the display (e.g., many of the elements are crosses and many of them are red, but the element to be detected is the only red cross in the display). Treisman's interpretation of these findings is that when an element is distinguished by a unique feature, it is easily detected because it is the only element which is associated with the "map" for that feature. In contrast, when an element is distinguished by a combination of features, it is not uniquely associated with any given feature map, and thus it can only be distinguished by the visual system by fusing the information from various feature maps into a perceptual whole (i.e., an object) at a particular location through visual attention.

On Treisman's (1988, 1998; Treisman & Gelade, 1980) interpretation, visual attention is mainly a spatial integrator, which does little to explain findings showing that it is easier to attend to a single object than to divide attention between many. However, Treisman's theory, along with the other research discussed in this section, reminds us that the role of attention in perception cannot be ignored or underestimated. Likely, attention and perception are mutually dependent, at least to some degree. That is, visual perception and attention are likely interactive and concurrent processes. On this view, perceptual grouping is not pre-attentive. Rather, the cost of dividing attention between objects might reflect the cost of organizing a scene into objects and groups instead of, or in addition to, any costs due to switching attention between objects.

3.4 How Many Objects? One, Two or More?

As a consequence of not having an operational definition of objectness, there is a lack of principled, objective criteria for individuating objects in a display. In many experiments, this is not a serious issue because the displays are so simple that there is no ambiguity as to which elements should be taken to be objects. For instance, the displays used by Duncan (1984) consisted only of a box with a dashed line running through it. This display is fairly unambiguous.

There are, however, examples of research where the assumptions as to the number of objects being displayed or perceived by the observer are questionable. For instance, Treisman et al. (1983) constructed displays that consisted of a word and a box-shaped frame shown simultaneously. The frame had a gap in one of its sides. In one condition, the word and the frame did not overlap spatially; this configuration is most likely perceived as two objects. However, in the other condition, the word was displayed inside the frame. In this case, Treisman et al. assumed that the display was perceived as a single (albeit complex) object. In both conditions, the distance between the gap in the frame and the word was the same. Since reaction times were lowest in the “one object” condition, even though there was no change in the distance between the targets the participants had to judge, it was concluded that the overlapping display was indeed perceived as a single object, and that attention is more easily allocated to a single object than to two objects. This explanation depends on the assumption that a spatial or spotlight model would necessarily predict that reaction time would be a function of the distance between the targets. However, it could be the case that the distance between the centroids of the box and the word would determine reaction times under a spatial model (for more on this see Driver & Baylis 1993). Thus, the assumption that the overlapping display forms a single object also requires a particular reading of spatial models of visual attention. None of these assumptions is necessarily warranted.

There are cases where a configuration that appears to be made up of separate objects might be perceived as a single object with internal structure. Lavie and Driver (1996) used a display consisting of two dashed lines intersecting at their mid-points and forming an “X”. Each of these lines was displayed in a different colour (red and green). The working assumption was that each line would be perceived as an object in its own right, and that dividing attention between both lines would produce a performance cost that would override any costs due to spatial separation. This assumption was again supported by the reaction times of the participants performing the experimental task: comparing two elements located on the same line took less time than comparing two elements located on different lines. Furthermore, targets located on separate objects and far apart from each other were compared just as slowly (relative to the single-object condition) as targets on separate lines but close together. However, an experiment by Herdman, Jarmasz and Johannsdottir (2000) suggests that the dashed-line “X” display used by Lavie and Driver can be parsed as either one or two objects. This study compared performance on the target comparison task under two conditions. In the first condition, each line making up the “X” was a different colour, as in the study by Lavie and Driver. In the second, both lines were displayed in the same colour; that is, the whole “X” was monochromatic (in this case, black). In the two-colour condition, the results mirrored those of Lavie and Driver. In the monochromatic condition, targets on the same line were again compared the fastest, but targets distant from each other and on separate lines were responded to faster than targets on separate lines and near to one another. This is unlike Lavie and Driver’s findings, and the two-colour condition of Herdman et al. (2000), where target separation had no effect on reaction times. This suggests the possibility that under some circumstances the two lines are indeed perceived as two separate objects, while in other circumstances the two lines appear to be perceived as a single “X”.

Research on applying object-based models to HUD or HMD environments illustrates more vividly how assumptions about the number of objects or perceptual groups in a display can have a significant practical impact. The assumption that has driven much of this research is that there are two relevant perceptual groupings in a HUD: the near domain (the instrumentation) and the far domain (the scenery). This has been stated perhaps most clearly in McCann, Foyle, and Johnston (1993; see also Foyle, Sanford, & McCann, 1991, and Wickens, & Long, 1995 for similar explicit statements). Many of the performance problems associated with HUDs have been explained as difficulties in dividing attention between perceptual groups. For instance, Martin-Emerson and Wickens (1997) examined the difference in the use of HUDs versus HDDs across different visibility levels in terms of dividing attention between the near and far domains. As pilots came in for an approach to land under different visibility conditions they had to hold a stable altitude and control lateral and vertical tracking. Flight path guidance was superimposed onto the path for the HUD condition and located below the windshield for the HDD condition. The results showed that for the HUD condition pilots were faster to respond to events within the HUD display and to control altitude when under zero visibility as compared to full visibility conditions. In full visibility pilots attend to the far domain, thereby making it more difficult to control the altitude and respond to events occurring in the HUD. However, lateral and vertical tracking errors also decreased in the full visibility condition.

Martin-Emerson and Wickens interpreted these results as showing that pilots had difficulty dividing attention between two perceptual “objects,” that is, the near and the far domains. This is consistent with an object-based hypothesis. According to this line of reasoning, in the full visibility conditions, pilots are attending to both domains, as shown by the fact that their performance on tasks requiring information from the far domain improves when this domain is visible. Because the far domain is now drawing some attention away from the near domain, pilots’ performance in tasks which require information from the near domain (i.e., the HUD) is degraded in full visibility conditions relative to the zero visibility conditions.

However, there is an alternative explanation which also makes use of an object-based attention model. Because pilots had to maintain their altitude and respond to events within the near domain, it is unlikely that they switched attention completely toward the far domain. Indeed, it is possible that the pilots successfully integrated the HUD flight-path symbology with the external environment but experienced difficulty attending to other information located in the HUD. That is, given that the near domain is itself made up of various symbols, the pilots might simply have experienced difficulty in integrating information within a single domain, rather than difficulty in dividing attention between domains.

As discussed above, it has been suggested that conformal and scene-linked symbology might be used to reduce cognitive tunnelling by fusing the near and far domains into a single perceptual group. Foyle, Sanford and McCann (1991), McCann and Foyle (1994; 1995) and Shelden, Foyle and McCann (1997) have studied the effects of using scene-linked symbology in a HUD environment. Although the results are promising, there were a number of methodological problems with these experiments. In particular, Foyle et al. (1991) found that adding elements onto the flight path, whether those elements are virtual buildings or numbers, increases the number of cues that the pilot can use to control the aircraft's flight path, thus making control of the aircraft less "inference-laden."[10] Accordingly, the enhancement in timesharing the flight-path task and the symbology-based tasks may not be due to a reduced requirement to switch attention across domains. Instead, this enhanced dual-task performance may be attributed to the reduced load associated with controlling the aircraft's flight path when more path cues are present.[11]

In another experiment, Martin-Emerson and Wickens (1997) tested the difference between conformal and non-conformal HUDs using symbology that differed only in terms of path guidance information. Both conditions included non-conformal symbology such as a vertical speed indicator (VSI), heading, speed, and distance. In the conformal condition, a virtual runway overlaying the actual runway provided path guidance. In the non-conformal condition, path guidance was represented by a localizer and a glide slope, a fixed aircraft symbol and a reference line. The participants' task was to approach to land under different visibility conditions. The results showed that for the non-conformal condition there was large variance in lateral tracking errors depending on visibility. In comparison, when pilots used the conformal symbology the lateral tracking errors were undifferentiated across the different levels of visibility, thus suggesting that conformal symbology helps pilots perform the task equally well in all visibility conditions.

---

[10] "Inference-laden" is not meant in quite the same sense as Sellars (1956), who used the term to indicate the pervasive presence of inference in perception, but rather to indicate that certain tasks require more conscious mental processing than others.

[11] As discussed on p. 190, Shelden et al. (1997) found that the advantage for scene-linked path markers obtains only when the path markers are relevant to the task, thus underscoring the difficulties in studying the effects of scene-linked symbology experimentally.

Interestingly, the condition of symbology type had no effect on vertical tracking errors; this is not accounted for in Martin-Emerson and Wickens (1997).

Martin-Emerson and Wickens (1997) concluded that the improvements in pilot performance were due to the conformality of the symbology. However, the object-based attention hypothesis can be applied to the Martin-Emerson and Wickens results although not necessarily under the assumption that conformal symbology fuses the two domains into a single perceptual group. For example, it could be argued that conformal symbology leads to better performance because the symbology forms a coherent object in and of itself. In accord with the object-based attention hypothesis, a sense of objectness presumably makes it easier for pilots to attend and process information within the conformal symbology. That is, the advantage of conformal symbology may not be due to the fact that the symbology is integrated into the external scene, as was assumed by the authors, but instead to the fact that the conformal symbology used made the runway a more easily processed perceptual object.

3.5 The Confounding of Spatial and Object-Based Factors

McCann, Foyle and Johnston's (1993) finding that responses to targets were significantly delayed when a cue was presented in the nontarget domain could be interpreted as showing that near (HUD) and far (external scene) domains form separate visual objects. In accord with the object-based attention hypothesis, target responses are slower when the cue occurs in the nontarget domain because it takes time to switch attention from one object (domain) to the other. However, a careful review of the

McCann, Foyle and Johnston procedure leads to an alternative explanation. In this experiment, pilots were required to perform an approach to land. As they approached landing, a three-letter cue (“IFR” or “VFR”) appeared either on the HUD or on the runway (external scene). The cue indicated where to look for a target among several geometric symbols that appeared on both the HUD symbology set and the runway. “IFR” (for instrument flight rules) indicated that the set of symbols on the HUD was relevant. “VFR” (for visual flight rules) indicated that the set of symbols located on the runway was relevant. The pilots were required to identify whether one of the symbols (the target) was a diamond or a stop sign: A landing was allowed only if the target was a diamond. Four boxes were located on the HUD to flank either side of the runway and another four boxes were superimposed onto the far domain in a similar position. The distance between the boxes was equal for both domains. The three geometric symbols appeared in three of the boxes on each domain and the cue would appear in the fourth box on either the near or the far domain. If the cue appeared on the HUD it filled the box in either the bottom left or the bottom right corner. If the cue appeared on the runway it filled either the top left or the top right box.

The results showed that participants were significantly slower in responding to the relevant target when the target and the cue where located on different domains. For example, when the “IFR” (indicating that target on the HUD is the relevant one) appeared in the display, pilots were faster to respond to the target when it was located on the HUD than on the runway.

As noted above, the object-based attention interpretation of the McCann, Foyle and Johnston (1993) result is that the near and far domains form separate visual objects: It takes time to switch attention from one object to the other. However, another plausible interpretation is that the slower responses in the cross-domain condition are due to there being two different types of spatial cueing. The sudden onset of the three letters is a form of exogenous or direct cueing that immediately draws attention to that location. In contrast, the interpretation of the three letters is a form of symbolic, or endogenous, cueing: the participants had to interpret the meaning of the three letters to determine the relevant target location. Attentional allocation is much slower with endogenous cues than with exogenous cues. Whereas attentional allocation can occur within 100 ms with exogenous cues (Wright & Ward, 1998), the allocation of attention with endogenous cues can require 300 ms and longer (Stelmach el al., 1997; Wright & Ward, 1998). On this view, when “IFR” was shown on the HUD, the pilots would be able to determine almost immediately whether the target on the HUD was a diamond or stop sign. There would be no need to interpret the cue itself. Accordingly, when the symbolic cue concurred with the direct cue, pilots were fast to respond. When the symbolic cue conflicted with the direct cue (a different location of the relevant target was indicated versus the direct cue), then responses were slow.

Further doubt is cast on the conclusion reached by McCann, Foyle and Johnson (1993) by the fact that it cannot be assumed *a priori* that the HUD and the outside scene each form distinct and coherent perceptual units. This being the case, the results found by

McCann et al. (1993) could be consistent both with limitations on object-based attention and with a spatial cue conflict.

A second, and related, problem in differentiating between spatial and object-based factors concerns the size of the display. In Duncan’s (1984) experiment the display was less than 1°. This allows for only a small and potentially insignificant variation in spatial distances. It could therefore be argued that object-based factors provide the only basis for selection, resulting in what seems like a general object-based attentional mechanism. Similarly with the paradigm used by Treisman et al. (1983), the spatial area relevant when the word was presented within the frame was much smaller than the area relevant when the word and the frame were presented separately. Therefore, the Treisman et al. results might have reflected the benefit of spatial proximity. Also, the onset of the frame may have exogenously grabbed attention making it difficult to read the word in the condition where the frame and word were presented separately.

Spatial factors other than location or separation between targets are also frequently confounded with object-based factors. Targets are often so different that the same-object effect can be the result of some experimental artefact (e.g., spatial frequency differences). In Duncan’s (1984) experiments, the two attributes of the line are available at a high spatial frequency whereas the two attributes of the box are available at low spatial frequency. The results may therefore reflect difficulties in processing or attending to different spatial frequencies. Furthermore, the targets used and the instructions given often indicate objects prior to the actual task. To wit, Duncan’s procedure in which participants are required to identify the height of a box and the texture of a line may bias participants toward processing the stimuli as objects.

To summarize, results from both basic and applied research on visual attention show that object-based and spatial factors are often confounded. This is largely due to the fact that the models of attention are not mutually exclusive, unlike what is sometimes implied in the experimental literature. Rather, they likely reflect two parallel and complementary systems. Evidence from neuroscience shows that there are two main pathways, one for spatial processing and one for object-based processing, in the visual system (Mishkin, Ungerleider, & Macko, 1983). Further, there are models of attention that integrate both spatial and object-based operators of attention (e.g., Logan, 1996). Clearly separating these two attentional systems in an experimental setting requires valid and reliable operational definitions of objects and perceptual groups, which are still lacking.

3.6 Interim Summary II: Problems Facing Object-Based Attention

The object-based theory of visual attention, as most commonly formulated, relies on the assumption that the representational structures selected by attention (a-objects) are governed by the Gestalt principles of perceptual organization. As noted above, there are three main problems with the Gestalt approach to perceptual organization: the confounding of conscious perception and attentional selection, the lack of criteria for individuating perceptual objects, and the lack of criteria to distinguish spatially-sensitive and spatially-invariant perceptual organization. In combination with certain assumptions about cognitive architecture and the correspondence between visual representations and objects in the world, the concerns with Gestalt perceptual grouping lead to three major problems with the object-based theory: the unwarranted assumption that perceptual grouping precedes attentional selection, a lack of criteria for determining how many objects are available for attentional selection by an observer, and the confounding of spatial and object-based factors in the allocation of attention.

It should be noted that, despite these problems, the evidence in favour visual attention being mediated by perceptual organization is strong. Interestingly, some of the strongest evidence for the object-based aspects of attention comes from research using motion. As discussed above, Driver and Baylis (1989) have shown that moving stimuli can hinder performance in an object-recognition task more than static stimuli that are near to the target stimuli. Also, Pylyshyn and Storm (1988) provided evidence that the tracking of multiple objects cannot be accounted for by spotlight models.

There are two reasons why motion might provide such robust evidence in favour of object-based attentional mechanisms. The first is that motion is an important part of object perception. Spelke, Gutheil and Van de Walle (1995) have shown that motion plays a crucial role in the development of object perception in infants. Furthermore, Wallis (2002) showed that the spatial continuity of the motion of objects was an essential factor in the establishment of object representations in memory. From this it has been assumed that elements are grouped into objects and coherent groups by common motion (Valdes-Sosa, Cobo & Pinilla, 1998; Yantis, 1998), a principle known as “common fate” in Gestalt psychology. On this view, elements moving together are grouped together, and static elements, with null common motion, are also grouped together.

The second reason for the importance of motion in object-based attention is that motion provides a means of isolating spatial and object-based factors in attention. Tipper and Weaver (1998) and Gibson and Egeth (1994) have conducted studies where the inhibition of return (IOR) of attention was studied with moving stimuli. As noted above, Tipper and Weaver cued attention to both moving objects and static locations in the display through which the objects passed, and measured the degree of IOR to the cued locations. They found that IOR was strongest to the cued objects; they also observed a weaker IOR effect to cued spatial locations. Gibson and Egeth used a similar paradigm, involving cue locations that were invariant with respect to a world-based frame of reference but which corresponded to different locations on the same object (achieved by rotating the object). They also found evidence for both spatially-referenced and object-based IOR.

Together, the studies on the role of motion in object perception and the use of motion in the study of IOR suggest that the use of motion could play a role in overcoming some of the difficulties facing research on object-based attention. Motion can be used to experimentally distinguish between spatial and object-based factors in attention, and could be used to explore the perceptual mechanisms that influence the organization of visual stimuli into objects that can be selected by attention. In doing so, the use of motion can provide a more secure basis from which to examine the temporal relationship between attentional selection and perceptual organization. These considerations motivated the experimental research presented in Chapter 4, which involved the study of attentional allocation within the context of grouping by motion.

4. EXPERIMENTAL WORK

4.1 Background

The present series of experiments was first conceived within a standard object-based attention framework. By the time the series was complete, the underlying framework had shifted away from a pure object-based model to one where action and observer intentions also played a role – the conative model, which will be discussed in Chapter 5.

The original goal of the experiments had been to extend a standard, static, object-based attention paradigm (Lavie & Driver, 1996) to dynamic displays. The Lavie and Driver paradigm had been designed to avoid what were considered shortcomings in much of the object-based attention literature of the time. The issues had to do mainly with the confounding of spatial and object-based factors, on the one hand, and the impact of task demands on the deployment of attention, on the other. Lavie and Driver argued that by presenting stimuli across only a limited field of view (i.e., only 1 or 2 degrees of visual angle), many experiments precluded the possibility of determining whether attention displayed object-based properties simply because the display was too small for selections on the basis of spatial location to be effective. As for task demands, they were concerned that either the instructions given participants (i.e., "compare the features in the two objects you see in the display") or properties of the task itself (such as requiring a global judgement about an object in order to compare the features of two objects) would bias participants' responses towards outcomes favourable to an object-based interpretation (or away, as the case may be), thus confounding the effects of attention and of task demands on performance.

Lavie and Driver (1996) suggested five criteria to avoid the problems outlined above: (1) the spatial separation of target elements should be manipulated across a wide extent (e.g., a few degrees of visual angle); (2) the target elements should have equivalent stimulus properties (e.g., texture, size) to rule out between-object effects due to extraneous factors such as spatial frequency; (3) the target elements should naturally be perceived as parts of objects without depending on the global properties of the objects for their identification; (4) the eccentricity of target elements should be equivalent; and (5) the instructions and task demands should bias participants to attend neither to particular locations nor to specific objects in the display. The display they used in their study, which conformed to the five criteria above, consisted of two lines arranged in a horizontal “**X**” that subtended about 13º of visual angle. Each line was made up of dashes about 1° of visual angle in length, and was presumably perceived as an object according to the Gestalt principles of collinearity (each line was made up of collinear dashes of equal length) and grouping by colour similarity (each dashed line had a distinct colour). The target elements were either a dot (i.e., a shortened dash replacing the usual 1° dash) or a gap (i.e., a dash replaced by an empty space), which according to the authors naturally looked as thought they were a part of the larger dashed line. These target elements appeared either close to each other but on different lines (the ‘near’ condition), far from each other (at different ends of the display) and on separate lines (the ‘far’ condition) or far from each other but on the same line (the ‘object’ condition). The participants were instructed to judge whether the target elements appearing in any given display were the same; the task did therefore not require the participants to make any judgements intrinsically based on spatial location or on the global properties of the dashed lines.

The two-dashed-line display developed by Lavie and Driver (1996) resulted in participants performing the judgement task more quickly when both targets appeared on the same line. Spatial factors were ruled out because target similarity in the ‘object’ condition, where the targets were separated by 8º of visual angle, was assessed more rapidly than in the ‘near’ condition, where targets were 3.1º apart. As judgements in the object condition were quicker than in the far condition as well, even though the distances between targets were comparable in both conditions, it was concluded that attention had been deployed in an object-based manner, i.e., participants’ judgement of the similarity of two targets were facilitated when the targets appeared within a single object, because attention is allocated to single objects rather than to spatially proximate elements. These experimental results seemed to strengthen the hypothesis that attention is object-based because the stimuli used avoided the methodological difficulties discussed by Lavie and Driver.

In parallel with the fundamental research on object-based attention, a number of researchers in the applied field of aviation psychology were making use of the object-based thesis to explain and propose solutions to the cognitive tunnelling problem that occurs with the use of HUDs. As was discussed in Chapter 2, the attention of a HUD user seems to be captured or ‘tunnelled’ by the HUD under certain conditions, to the detriment of information in the outside scene. Most research on cognitive tunnelling is premised on the following assumptions: attention is object-based, and the HUD and the outside scene each form a perceptual group according to the Gestalt principles of similarity of colour and common fate. On this line of reasoning, it would be more effortful to switch attention between two objects than to attend to only one, and if an event occurs in one ‘object’ while a pilot’s attention is on the other, or even when it is divided between both, a pilot’s ability to detect or identify the event might be compromised. A corollary to this is that if elements of the HUD could be perceptually ‘fused’ with the outside scene, using conformal symbology (Wickens & Long, 1995) or scene-linked symbology (McCann & Foyle, 1995), cognitive tunnelling might be mitigated.

As discussed above, a number of HUD studies tend to support the object-based account of tunnelling as well as the use of conformal symbology to relieve tunnelling. However, this research seems to take it for granted that HUDs and the external scene are perceived as single, distinct objects. There is, in fact, no evidence to confirm this specific hypothesis, rather than the more general claim that pilots can attend more effectively to a single source of information than to information distributed across the pilots’ visual domains. The claim that a HUD constitutes a single perceptual object crucially depends on whether common fate and similarity of colour are in fact sufficient for Gestalt-type perceptual grouping to occur. To my knowledge, these claims have not been explicitly tested, especially not in a display where the elements to be grouped are superimposed over (not just displayed alongside) other elements, and where the other elements might themselves be moving, much as what is seen in an actual HUD-equipped cockpit. Thus, it seemed natural to try to extend the Lavie and Driver paradigm to a dynamic display that would allow us to test the claim that common fate could support the perceptual grouping of a HUD into a single object. Note that the role of colour in HUD perception is not addressed in the following experiments. However, data from the ACE Lab at Carleton University (Herdman, Jarmasz, & Johannsdottir, 2000) show that while colour plays an auxiliary role in facilitating grouping, it is not necessary for grouping.

An experimental display was thus devised to both extend Lavie and Driver's (1996) static paradigm to dynamic contexts, both out of theoretical interest and to examine the claim that object-based attention could in fact be a factor underlying cognitive tunnelling with HUDs. The display was to achieve this by complying with Lavie and Driver's five criteria reviewed above while using common fate to group elements into objects. This defined the general logic of the experiments reported below. That is, every experiment reported in this research required participants to perform a two-alternative forced-choice (2AFC) task on the features of elements of groups which were themselves defined by some variation of the principle of common fate. The displays, which consisted of two groups of (either 3 or 4) identically coloured dots where at least one of the groups was moved through an elliptical trajectory on a computer monitor, subtended large areas of visual space (8 to 13º). Thus, in most experiments (except Experiment 4), common fate was the only stimulus-dependent factor distinguishing both groups[12], thus ensuring that the effects of common fate were isolated. The trajectories of the groups were such that both groups were overlapping for large proportions of the duration of each experimental trial, and the positions of individual dots within a group were randomized on each trial, ensuring that memory of target locations played no role in attentional selection. In each trial, two dots (randomly selected by the computer program to be either both within one of the groups or one in each group) would simultaneously change colour (either both to the same colour or each to a different colour). The task required participants to judge whether the target dots were the same colour, and thus did not require global judgements about the groups of dots.

Standard formulations of the object-based attention hypothesis lead to the prediction that attention will be most easily allocated to sets of stimuli that form perceptual wholes according to the Gestalt principles. Thus the critical hypothesis in every case was that target dots appearing within a single group would be judged more rapidly than targets spread across both groups. However, existing object-based accounts do not address whether all grouping factors (Gestalt or not) are equally important in the formation of wholes, and whether and how object-based factors (i.e., factors related to perceptual organization) interact with other factors which might affect attention allocation, viz. spatial factors, feature-based factors, and task-related factors. In the series of experiments described below, various factors (attentional focus, connectedness of dots,

---

[12] Technically speaking, in Experiments 1 through 4, the groups were also distinguished by number of dots (3 vs. 4). However, no effects consistent with dot number differences were found. On a theoretical level, studies on subitizing suggest that groups of 3 or 4 dots not would be attended to much differently (review in Pylyshyn et al., 1994).

type of motion) were manipulated to examine whether grouping by common fate would be sufficient in and of itself to produce the object effect unconditionally. On the whole, the critical hypothesis – that targets within a single group defined by common fate would be compared more quickly than targets distributed across groups – was confirmed. However, what emerged with growing clarity from the experiments was that the more relevant an experimental factor was to the 2AFC task, the more it interacted with grouping by common fate to affect the deployment of attention. In fact, common fate by itself was revealed to be a relatively weak factor in the allocation of attention to portions of a display (in terms of magnitude of effect), whereas the task-relevance of a factor generally had a greater influence. This has significant implications for the object-based hypothesis – namely, that the object-based aspect of visual attention manifests itself most strongly under particular conditions, especially when the object-related aspects of a display are relevant to a task. That is, object-based attention is not an unconditional, invariant property of visual attention, but rather attention becomes more-or-less object based depending on various contextual factors. The theoretical implications of the experimental results of the present research will be discussed more fully following a description of the experiments themselves.

4.2 Experiment 1

Experiment 1 was designed to examine whether grouping by common fate alone would produce the object effect as observed in most object-based experiments using static displays (e.g., Duncan, 1984; Treisman et al., 1983; Lavie & Driver, 1996). It was also designed to conform to the five criteria for experiments on object-based attention proposed by Lavie and Driver (1996) and discussed above. To this end, as detailed below, the experimental display was designed such that common fate provided the only basis for perceptually distinguishing the perceptual groups (objects) in the display, and that the experimental task neither required information about the global properties of the objects nor involved elements which were not perspicuously components of the objects formed by common fate. The display consisted of identical grey dots differentiated by common fate into one group of static dots and one group moving in unison along an elliptical trajectory. The central hypothesis in the experiment was that participants would perform the colour comparison more quickly and more accurately when the target colour change occurred for dots within a single group than for dots distributed across both groups. No hypothesis was made as to whether either of the groups (static or moving) would be favoured by the participants. Thus, the critical comparison was between latencies and accuracy rates for targets appearing across both groups versus targets both appearing in any one of the groups. Since the distance between the target dots was randomly determined in each trial, the null hypothesis was that target location (i.e., within a particular group, or both) would have no effect on latencies or accuracy. However, space-based accounts provided the rival hypothesis that latencies and accuracy would be positively and negatively correlated, respectively, with distance between targets.

4.2.1 Method

Participants

Eleven participants (7 female, 4 male) volunteered and were paid a nominal fee ($10 Canadian) to participate in the experiment. All participants had normal or corrected visual acuity, colour vision, and contrast sensitivity, were drawn from the Carleton University community, and ranged in age from 20 to 45 years.

Apparatus and Stimuli

The visual display was presented on a high-resolution Hitachi 21 inch colour monitor. The display was controlled by a Cambridge Research Systems VSG 2/3 video board installed in a Pentium-powered IBM compatible computer. Responses were collected at <1 ms accuracy using a micro-switch response box connected to the VSG board.

The display consisted of seven light grey dots shown against a dark grey background at random positions (Criterion 2, Lavie & Driver, 1996). The whole display subtended roughly 14° of visual angle and each dot had a diameter of 0.3° (Criterion 1, Lavie & Driver, 1996). Four of the dots stayed stationary during a trial, whereas the other three dots moved in unison through an elliptical trajectory. The trajectory's parameters (direction of rotation, height and width) were randomly varied from trial to trial (Criteria 2, 3 and 4). The linear velocity of the moving dots ranged from 3° per second to 7° per second. During the trial, two of the dots (targets) in the display changed to either of two possible colours, green or red. The two target dots were randomly assigned to appear both in the static group, both in the moving group, or one in each group, each outcome being equiprobable (probability 1/3). The target dots were also randomly assigned either to each change to a different colour (probability ½) or to both change to the same colour (red, probability ¼, or green, probability ¼). The parameters of the moving group's trajectory, the starting position of the dots, and the location and colour of the target dots were varied completely independently of one another (thus, the fact that the static group had 4 dots did not make it more likely that that group would contain one or both of the target dots). The display gave no indication whatsoever as to which of the dots in the display would turn out to be the target dots.

Design and Procedure

A two-way repeated-measures design was used with three levels of target location (static object, moving object, across objects) and two levels of target colour (same or different), for a total of six distinct and equiprobable conditions. On each trial, the dots were displayed for a variable interval lasting between 2 s and 6 s, during which time all dots remained light grey and the moving dots described their trajectory. At the end of the variable interval, two randomly selected target dots changed colour. The static and moving dots remained visible (including the target dots) until the participant responded, or until the display timed out after three seconds. Participants were to indicate as quickly and accurately as possible whether the two target dots were the same or different colour. Responses were indicated with the preferred hand by pressing one button on the response box for a "same" judgment, and another button for a "different" judgment. Participants were instructed to distribute their attention to the display as a whole and to avoid focusing attention on either the static or the moving dots (Criterion 5, Lavie & Driver, 1996).

Each participant completed 8 blocks of 60 trials. The first block for each participant was a practice block, leaving a total of 420 experimental trials per participant. The participant initiated each block, whereas each trial was initiated automatically following a 2-second pause.

Statistical Analysis

The following measures were collected for each participant on each trial: reaction times (RTs), counted from the instant the target dots changed colour to the instant the participant's button press was detected; accuracy of the response in the 2AFC task; the length of the interval before the target dots changed colour; distance between the target dots at the instant they changed colour; and distance between the target dots at the instant the participant's response was detected. In order to reject the null hypothesis, only analysis of the reaction times and accuracy rates per participant was required. However, to rule out the spatial hypothesis, analysis of RTs and accuracy as a function of spatial location was also performed.

The accuracy rates per participant were analysed with a 3 (target location) by 2(target colour) repeated-measures analysis of variance (ANOVA). The median RTs on correct trials per participant were also subjected to repeated measures ANOVAs. Median reaction times were used to diminish the effects of outliers without sacrificing data points to an arbitrary cut-off criterion (e.g., the 2.5 standard deviation rule commonly used in reaction time research). The effects of distance on RTs and accuracy were analyzed by performing bivariate correlations between either distance at target onset or distance at response and reaction time.

Pairwise comparisons were conducted using the within-subject confidence interval technique developed by Loftus and Masson (1994). This method allows for rapid and reliable pairwise comparisons for any two conditions in a repeated-measures design (be it factorial or one-way, multi-level) without having to re-compute the comparison every time. The mathematics behind the method produce a visual heuristic for judging significance: if the overlap between the confidence intervals of two means is less than ¼ of the whole interval (i.e., less than one-half of the distance between the mean and the upper or lower bound of the interval), the two means are significantly different. It should be noted that whereas a standard confidence interval (based on the between-subjects variance) can be used to estimate the value of a sample statistic, the within-subjects confidence interval (based on the within-subjects variance) is best conceptualized as a tool for estimating of the variance in the difference between two cells in a repeated-measures design, with no bearing on the statistical validity of the estimate of any particular sample statistic in the design.

4.2.2 Results

Accuracy Rates

Accuracy rates were generally high (average of 95%). An ANOVA revealed no significant differences between accuracy rates in each of the conditions defined by the combinatorial combination of target locations and target colour.

Reaction Times

RTs on correct trials were first analyzed with a 3(target location) by 2(target colour) repeated-measures ANOVA. The analysis revealed a significant main effect of target location, $F(2,20) = 5.465$, $p < .05$, $MSE = 340.273$. No other main effect or interaction achieved significance. The RTs were therefore subjected to a one-way, three-level (target location only) ANOVA, which confirmed the effect of target location found in the 2-way ANOVA, $F(2,20) = 5.735$, $p < .05$, $MSE = 165.205$.

A pairwise comparison of the RTs across target location conditions was performed using the within-subject confidence interval method (Loftus & Masson, 1994). The results of the analysis are displayed in Figure 6. They clearly show that reaction times are slowest when each target appears in a different group, and are fastest (by an average of 14 ms) when the targets both appear in a single group. The confidence intervals reveal no significant difference between reaction times for targets within the static group and targets in the moving group.

Reaction Times as a function of target distance

A bivariate Pearson product-moment correlation between RTs for correct trials and distance at target onset produced only a very weak correlation ($r<.029$) which failed to reach significance. A separate analysis of the same correlation for each target location condition (across layers, static layer, moving layer) also revealed only weak correlations ($r = .002$, $r = .033$, $r = .048$ respectively) which also all failed to achieve significance.

A bivariate correlation was then performed between RTs for correct trials and target distance at response. This analysis revealed a significant ($p < .01$) but very weak positive correlation between distance at response and reaction time (r = .051, accounting for only 0.26% of the variance). An analysis by target layer condition revealed a significant correlation only for targets displayed across groups (r = .061, $p < .05$, accounting for only 0.37% of the reaction time variance). Thus RTs were not meaningfully correlated with target distance.

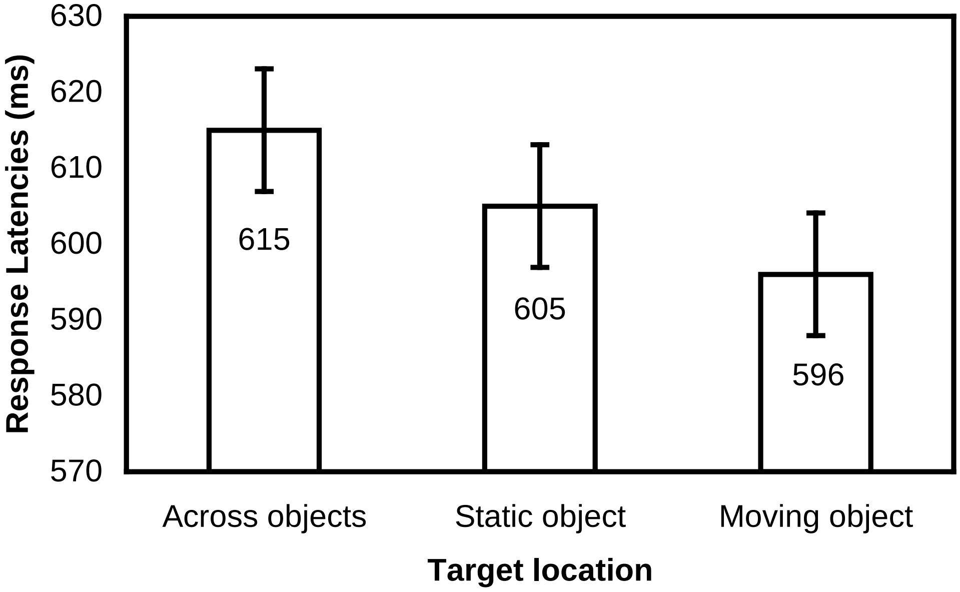

Figure 6: Response latencies (ms) and 95% within-subject confidence intervals (8.08 ms) for Experiment 1. Across objects: one target appeared on the static object and one target on the moving object. Static object: both targets appeared on the static object. Moving object: both targets appeared on the moving object.

4.2.3 Discussion

In this experiment, participants performed the 2AFC task most rapidly when both targets appeared within a single group, and most slowly when the targets appeared each in a different group. Furthermore, there was no relationship to speak of between the distance between the targets and response latencies. The object-based hypothesis, that features within a same object would be compared more effectively than features across objects, was thus supported. Also, the rival spatial hypothesis, that differences in responses between target location conditions might be due to spatial separation of the targets being larger across than within groups, was ruled out by the lack of correlations between response latencies and target separation.

Experiment 1 demonstrates that object-based effects also obtain when the objects in question are defined by common fate – common motion or lack thereof. This extends the fundamental result in the object-based literature, namely that the features of a single static object can be attended to more readily than features from multiple objects. It also extends the results of Driver and Baylis (1989), who found that grouping by common fate can have a bigger influence than grouping by spatial proximity in the flanker task (Eriksen & Eriksen, 1974) discussed in Chapter 2. In that study, common fate was shown to interfere with a judgement about a single feature within a single object, whereas the results of Experiment 1 reported above also suggest common motion can play a role in the segregation of a display into perceptual groups that can be selected by attention. This is in line with a rich literature that suggests that object motion is a strong determinant of object perception per se (Spelke et al., 1995; Wallis, 2002).

Experiment 1 also provides a necessary first step to establishing (rather than assuming) a link between object-based attention and cognitive tunnelling in HUDs. The elements of a HUD have an overall common motion relative to the outside scene that presumably helps to perceptually differentiate it from the outside scene. The results reported above provide indirect support for the claim that common motion of the type that distinguishes HUD from scenery can indeed produce object-based attentional effects. Without this, the claim that cognitive tunnelling occurs because of the object-based nature of visual attention would be much more difficult to make, as it would rely solely on the grouping of the HUD into a single object based on similarity of colour. This argument could only be made for monochrome HUDs (many HUDs still are rendered by green raster technology), and even in those cases, there is the added difficulty that colour is not a very effective grouping principle (e.g., see Folk & Remington, 1999; Herdman et al., 2000). An object-based account of tunnelling is tenuous at best without positive evidence that common fate alone can produce object-based effects.

The use of HUDs, however, raises an important issue for visual attention: whereas in Experiment 1 participants were instructed to attend to the display in a global manner, not favouring any particular region or object, a HUD user typically attends preferentially to one or another part of the outside scene or the HUD as the current task requires. Wickens and Long (1995) exploited this fact by having unexpected “far domain” events (a runway incursion) occur right after pilots’ attention had been forced onto the HUD by blocking out the outside scene with thick fog. This procedure led to dramatic effects (either a complete failure to notice the runway incursion, or a long delay to detect it, on the order of 3 to 4 seconds). It could be argued that this is an extreme case of a phenomenon which commonly occurs with HUD use, namely, the HUD user must have some awareness of elements in one object while having their attention preferentially drawn to another object – that is, the pilot must asymmetrically divide visual attention between the HUD and other elements. Experiment 2 was designed to shed some light on this aspect of object-based attention.

4.3 Experiment 2

In simulator-based studies of cognitive tunnelling, as presumably in real-life use of HUDs, pilots must divide their attention between the HUD and the outside scene in an asymmetrical fashion: they must focus their attention on the HUD versus the external scene as a function of task demands, while nevertheless still monitoring the other "domain" of their visual world. Wickens (Wickens and Long, 1994; see also Fadden, Ververs & Wickens, 2001) suggests that under extreme conditions, this asymmetrical division of attention can produce cognitive tunnelling to the point that certain perceptually salient elements of the display are not perceived. Experiment 2 examined whether objects that are differentiated through common fate can indeed be selectively attended in this asymmetrical fashion. That is, the critical hypothesis in Experiment 2 was that attending preferentially one object would improve performance on the 2AFC task for targets appearing within the attended object relative to targets appearing either in the unattended object or across objects. No hypothesis was made as to differences in performance when targets appeared in the unattended object versus across objects. The display and procedure were identical to that of Experiment 1, except that participants were now instructed to focus their attention either on the static or the moving group of dots. The null hypothesis in this case would be that focusing attention on a particular group will not have an effect – the results would replicate those of Experiment 1. A spatial hypothesis would be confirmed, as in Experiment 1, only if a relationship between target separation and response latencies were demonstrated.

4.3.1 Method

Participants

Six volunteers (3 male, 3 female) completed the experiment. All participants had normal or corrected-to-normal visual acuity, colour vision, and contrast sensitivity, were drawn from the Carleton University community, and ranged in age from 20 to 45 years. Three were male and three were female.

Apparatus, Stimuli, Design, Procedure and Statistical Analysis

The apparatus and stimuli were the same as in Experiment 1 except that a ViewSonic PS790 17 inch colour monitor was used. A repeated-measures design was used. There were 12 conditions in the experiment, which were defined by the 2 x 3 x 2 factorial combination of attentional focus (attend static-object vs. attend moving-object), target location (static object, moving object, across objects) and target colour (red vs. green).

The procedure was the same as that of Experiment 1 with the exception that each participant completed 6 blocks of 60 trials (total of 360 trials). For each block, the participant was instructed to attend to either the moving object or to the static object. The order of attentional focus was counterbalanced across participants. The same measures were collected and the same statistical methods were used as in Experiment 1.

4.3.2 Results

Accuracy Rates and Reaction Times

As in Experiment 1, accuracy rates were generally high (average of 93%), and the differences across conditions failed to achieve significance. They were therefore not further analyzed.

Mean latencies on correct trials were computed for the 12 experimental conditions from the median latencies for each participant. A 2(focus) by 2(target colour) by 3(target location) repeated-measures ANOVA revealed an interaction between focus and target location, $F(2,10) = 38.55$, $MSE = 7273.9$, $p < .01$, and a main effect of target location, $F(2,10) = 4.352$, $MSE = 3376.9$, $p < .05$. No other main effect or interaction achieved significance. In particular, since the effect of target colour was not significant, data from both target colour conditions were pooled in all subsequent analyses.

Within-subject confidence intervals using the interaction *MSE* term from the 2(focus) by 3(target location) ANOVA were used to perform pairwise comparisons between target locations. As shown in Figure 7, participants responded markedly faster when both targets appeared in the attended object. When attending to the static object (left panel) participants responded most quickly when both targets appeared in the static object (middle bar), as opposed to when both targets appeared in the moving object (right bar) or when the targets appeared across the two objects (left bar). In contrast, when attending to the moving object (right panel), participants responded most quickly when both targets appeared in the moving object (right bar) as opposed to when both targets appeared in the static object (middle bar) or when targets appeared across the two objects (left bar). This contrast accords with the significant target location by focus interaction found in the 2 by 2 by 3 ANOVA above. Note that the difference between response latencies to targets appearing both within the attended object and latencies to targets appearing either both within the unattended object or across objects was larger than the effect observed in Experiment 1, on the order of 80ms.

Reaction Times as a Function of Target Separation

Pearson correlations were computed between RTs and target separation at onset, and between RTs and target separation at response. The correlations were first computed for all data pooled across target location and focus conditions, then with data separated according to target location, then with data separated according to both target location and focus condition. The results are displayed in Table 1.

The positive correlations between RTs and target separation failed to achieve significance, and did not account for a meaningful amount of the RT variance. The spatial hypothesis was thus disconfirmed, as in Experiment 1. However, weak but significant *negative* correlations between RTs and target separation were found for targets appearing within a single group, regardless of the group (static or moving), or, more interestingly, regardless of focus condition. These results are the opposite of those predicted by the spatial hypothesis, and thus rule it out. The reasons for this pattern of correlations are unclear, and as the pattern was not seen in subsequent experiments, it is not discussed further.

Table 1: Correlations between reaction time and target separation in Experiment 2

	Separation at onset[a]	Separation at response[a]
All pooled	r = -.041	r = -.049*
Across layers		
pooled all focus	r = .085*	r = .071
focus static	r = .064	r = .073
focus moving	r = .111*	r = .067
Static layer		
pooled all focus	r = -.135**	r = -.135**
focus static	r = -.156**	r = -.156**
focus moving	r = -.104*	r = -.104*
Moving layer		
pooled all focus	r = -.158**	r = -.158**
focus static	r = -.169**	r = -.169**
focus moving	r = -.143*	r = -.143*

Notes:

* correlation significant at the .05 level (2-tailed)

** correlation significant at the .01 level (2-tailed)

[a] only the correlations between RTs and target separations for all layers pooled and across layers differ, because only in the across layers condition does target separation vary from onset to response

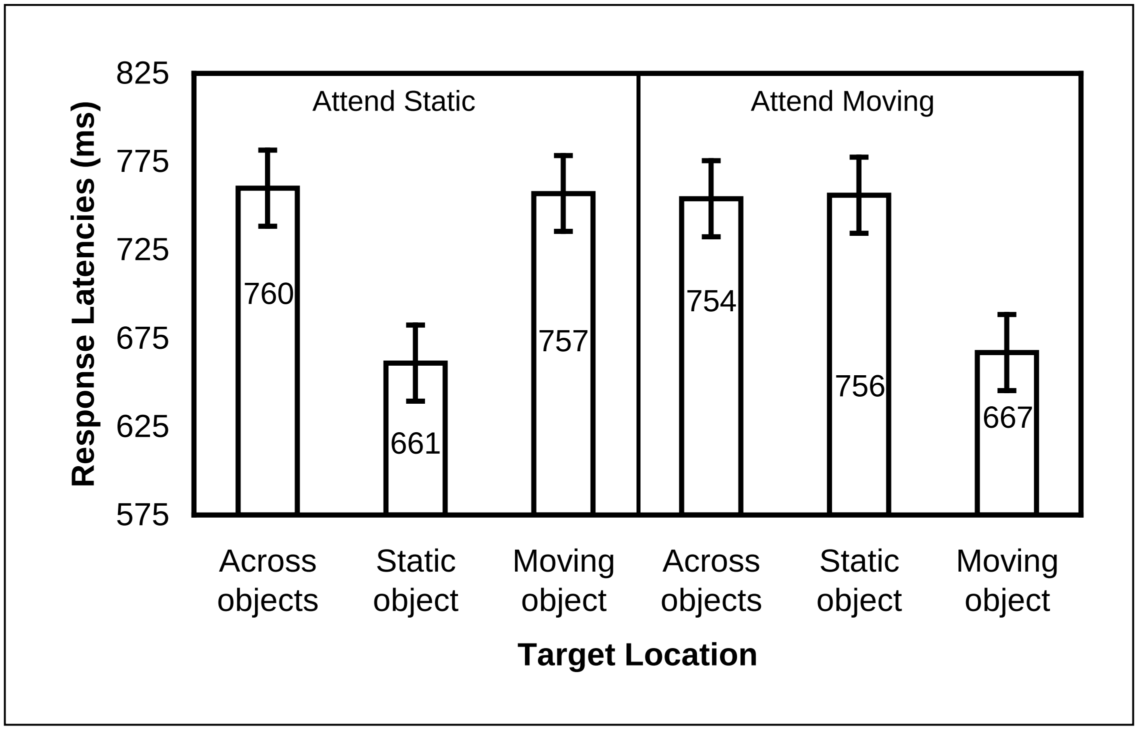

Figure 7: Response latencies (ms) and 95% confidence intervals (22.56 ms). Across objects: one target appeared on the static object and one target on the moving object. Static object: both targets appeared on the static object. Moving object: both targets appeared on the moving object. Left panel (Attend static) shows trials where attention was focused on the static. Right panel (Attend moving) shows trials where attention was focused on the moving object.

4.3.3 Discussion

The results for Experiment 2 support the original hypothesis, viz., that focusing attention on a given object will improve performance in the 2AFC task for features of the attended object relative to features of the unattended object, or features displayed across objects. Thus, deliberately focusing attention on an object results in an asymmetry in divided attention which, in extreme cases, might result in cognitive tunnelling as observed in HUDs. Relative to the tunnelling observed in HUDs, which generally involves unexpected events in the unattended domain, the results of Experiment 2 suggest that weaker but nevertheless significant asymmetries in divided attention might occur even with normal HUD usage, where the events in all domains are expected but task and/or contextual demands make one domain more "relevant" than the other.[13]

One limitation of this experiment is that the focusing of attention is the result of an explicit instruction to do so, rather than of an implicit task demand, and thus might not reflect normal HUD usage, or the way people normally attend to their environment. This issue is addressed to some extent by Experiments 6 and 7 below. Another limitation is that a direct comparison between the results of Experiments 1 and 2 is impossible, because the experiments involved different numbers of participants. Such a comparison, however, is desirable, as it would allow us to say whether focusing attention on an object

---

[13] Divided attention is generally discussed as though it were a binary phenomenon – either attention is divided, or it is not – though the experimental results from the vast literature on divided attention are most supportive of the claim that attention can be asymmetrically divided between relevant stimuli (see in particular Johnson & Yantis, 1995).

modulates object-based attention by inhibiting the processing of unattended stimuli, by enhancing the processing of attended stimuli, or both. The practical implications are enormous, because if focused attention does inhibit the processing of unattended stimuli, HUD designs or tasks which encourage the focusing of attention onto an object also carry a hidden cost, i.e., decreased performance for tasks involving unattended[14] but nevertheless important stimuli.

4.4 Experiment 3

Experiment 2 suggests that deliberately focusing attention on one object creates an attentional asymmetry – features of the attended object are processed more effectively than features located on objects that are not the centre of attentional focus. However, these results do not allow us to say whether the relative processing advantage for the attended object comes at the cost of inhibiting attention to other parts of the scene rather than enhancing processing in the attended object, or both. In parallel with these results, cognitive tunnelling research suggests that object-based attention is inhibitory insofar as attending to a HUD results in diminished processing of events in the (less-attended) external scene.

Taken together, the results from Experiment 2 and from cognitive tunnelling suggest that focused object-based attention inhibits attention to non- (or less-) attended

---

[14] ‘Unattended’ will be used throughout to refer to stimuli that are not at the focus of attention, with the understanding that so-called “unattended” stimuli can, and, I would argue, usually do have some attention directed onto them.

objects, and that cognitive tunnelling is an extreme case of this attentional asymmetry where all attentional resources are focused onto a single object, in this case the HUD. However, they are neutral as to whether focused attention can also lead to enhanced processing for an attended object. The present experiment was designed to directly examine the facilitory versus inhibitory nature of object-based attention by providing a baseline condition, in which attention was not deliberately focused on any object (analogous to Experiment 1), against which to compare the results of focusing attention on an object. Participants performed the same task as in Experiment 2, whereby attentional focus was varied from block to block. However, three rather than two focus conditions were used: participants were required to focus on the static object on one third of the trials, on the moving object for another third of the trials, and on neither object specifically ('no focus' condition) for the remaining third of the trials.[15] The primary hypothesis for this experiment, based on the results of Experiment 2 and of research on cognitive tunnelling, was that focusing attention would mainly have an inhibitory effect – that is, it was expected that features of an object would be processed more efficiently when the object was not the "centre of attention" than when no one object was the locus of attentional focus. Secondarily, it was hypothesized that features of an object would be processed more efficiently when the object was the locus of attentional focus than when

---

[15] The design of this experiment is effectively a combination of the designs of Experiments 1 and 2.

no object was attentionally “favoured”.[16] Finally, it was expected that the pattern of results within each attentional focus condition would be consistent with the previous two experiments (Experiment 1 for the results of the ‘no focus’ condition, and Experiment 2 for both of the conditions where attention was focused onto a particular object).

4.4.1 Method

Participants

Five participants (2 female, 3 male) with normal visual acuity, colour vision, and contrast sensitivity participated in the experiment. All were volunteers from the Carleton University community. Ages ranged from the mid-twenties to the mid-forties.

Apparatus, Stimuli, Design, and Procedure

The apparatus and stimuli were the same as in Experiment 2 with the exception that the targets were either blue or yellow. The colours were changed from green and red as it was suggested by a participant in a previous experiment that green and red have a strong cultural significance (traffic lights) and their use might confuse participants. There were nine conditions in the experiment, defined by the 3 x 3 factorial combination of target location (static object, moving object, across objects) and attentional focus (attend static-object vs. attend moving-object vs. no focus). All factors were manipulated within participants.

---

[16] This hypothesis is said to be secondary in that it was seen as less likely to be supported by experimental data than the primary hypothesis, based largely on HUD research an on a visual comparison of the data from Experiments 1 and 2.

The same procedure as in the previous experiments was used. Instructions were very similar to those in Experiment 2, with the difference that participants were given three attentional focus strategies (attend static-object, attend moving-object, no focus). For the ‘no focus’ condition, participants were instructed to spread their attention across the whole display, without focusing on either the moving or the static object. It should be understood that while participants were instructed to avoid focusing on a specific object, nothing in the display or the task compelled them to do so. However, the results of Experiment 2 suggest that participants are able to focus attention onto a moving group of dots, whereas Experiment 1 suggests that in the absence of instructions participants do not favour the static or the moving group. Thus, it is predicted that the instructions to focus or not focus attention will produce results analogous to those of Experiments 2 and 1, respectively.

Each participant completed 7 blocks of 60 trials. The first block constituted a practice block and did not figure in the analysis (a total of 360 trials per participant were analyzed). Attentional focus conditions were blocked within each participant.

4.4.2 Results and Discussion

As with the previous experiments, accuracy rates were high (average of 94%), did not differ significantly across conditions and thus are not discussed further. Mean latencies on correct trials were computed for the nine experimental conditions from the median latencies for each participant. Also, as reaction latency analyses in the previous experiments failed to reveal significant differences for response condition (i.e., same or different target colours), response condition was not considered in the following analyses.

There were no significant correlations between response times and target separation at target onset and at response.

A visual inspection of the data (see Figure 8) suggest that the results are consistent with the findings of Experiment 1 (latencies for the ‘no focus’ condition appear to be slightly faster for targets located within the same object) and replicate those of Experiment 2 (focusing on one of the groups of dots produces a pattern of reaction latencies wherein targets in the attended object are fastest, and targets in the unattended object and across objects are responded to more slowly). Figure 8 also suggests that the effect of focusing attention on one object consists mainly of a slowing down of responses to unattended objects, and that any enhancement of reaction times within the attended object (relative to targets in the same object in the ‘no focus’) is small, if any.

A 3(focus condition) by 3(target location) ANOVA of the latency data revealed only one significant effect for the interaction between focus and target location, $F(4,16) = 11.4$, $MSE = 272.4$, $p < .01$; no other effect achieved significance. This is consistent with the fact that RTs have very different patterns depending on focus condition, as seen in Figure 8. To determine whether the apparent differences in response latencies in the attend static and attend moving conditions were significant relative to the ‘no focus’ condition, pairwise comparisons using within-subject 95% confidence intervals (Loftus & Masson, 1994) based on the *MSE* from the 2-way interaction of the ANOVA above were performed. The confidence intervals are shown as bi-directional error bars around the data in Figure 8. As in Experiment 2, participants responded most quickly when both targets occurred in an attended object. When attending to the static object (open bars) participants responded quickest when both of the target dots appeared in the static object (Panel b), as opposed to when both targets appeared in the moving object (Panel c) or when the targets appeared across the two objects (Panel a). Similarly, when attending to the moving object (dark shaded bars), participants responded quickest when both targets appeared in the moving object (Panel c) as opposed to when both targets appeared in the static object (Panel b) or across objects (Panel c). For the 'no focus' condition (light shaded bars), target location did not significantly influence response latencies. This result does not strictly match that of Experiment 1, where the single-object conditions had the fastest response times. However, the pattern of reaction times for the 'no focus' condition in the present experiment is similar to that found in Experiment 1; therefore, given the lower power per condition in Experiment 3, the findings here can be said to be consistent with those of Experiment 1.

The facilitory versus inhibitory nature of object-based attention was examined by comparing latencies in the static-attend and moving-attend conditions to those in the 'no focus' condition. Of primary interest are the sets of trials where both targets appeared in the same object (static or moving). Evidence that object-based attention has facilitory effects was limited to the condition where participants attended to the moving object. As shown in Figure 8, focusing attention on the moving object facilitated processing of targets relative to the 'no focus' condition: participants responded significantly faster to targets that both appeared in the moving object when the moving object was attended (Panel c: dark shaded bar) as compared to the 'no focus' condition (Panel c: light shaded bar). In contrast, responses to targets that both appeared in the static object did not differ significantly when the static object was attended (Panel b: open bar) as compared to the ‘no focus’ condition (Panel b: light shaded bar). Thus, focusing attention on the static object did not significantly facilitate processing of targets relative to the ‘no focus’ condition.

Evidence for inhibition was found in both the static-attend and the moving-attend conditions. As shown in Figure 8, participants responded more slowly to targets that appeared in the static object when their attention was focused on the moving object (Panel b: dark shaded bar) as compared to the ‘no focus’ condition (Panel b: dark shaded bar). Similarly, participants responded more slowly to targets that appeared in the moving object when their attention was focused on the static object (Panel c: open bar) as compared to the ‘no focus’ condition (Panel c: light shaded bar).

In sum, Experiment 3 shows that focusing object-based attention on stimuli that are grouped by common fate provides some processing facilitation. However, this facilitation is not robust and it only occurs when attention is placed on a moving object. Importantly, object-based attention to stimuli that are grouped by common fate inhibits processing of information in unattended objects consistently, and to a larger degree than it facilitates processing within attended objects (30 vs 9 ms, respectively). These results are consistent with the notion that cognitive tunnelling reflects an extreme case of focusing of object-based attention onto a HUD at the cost of inhibiting processing of information in the external scene.

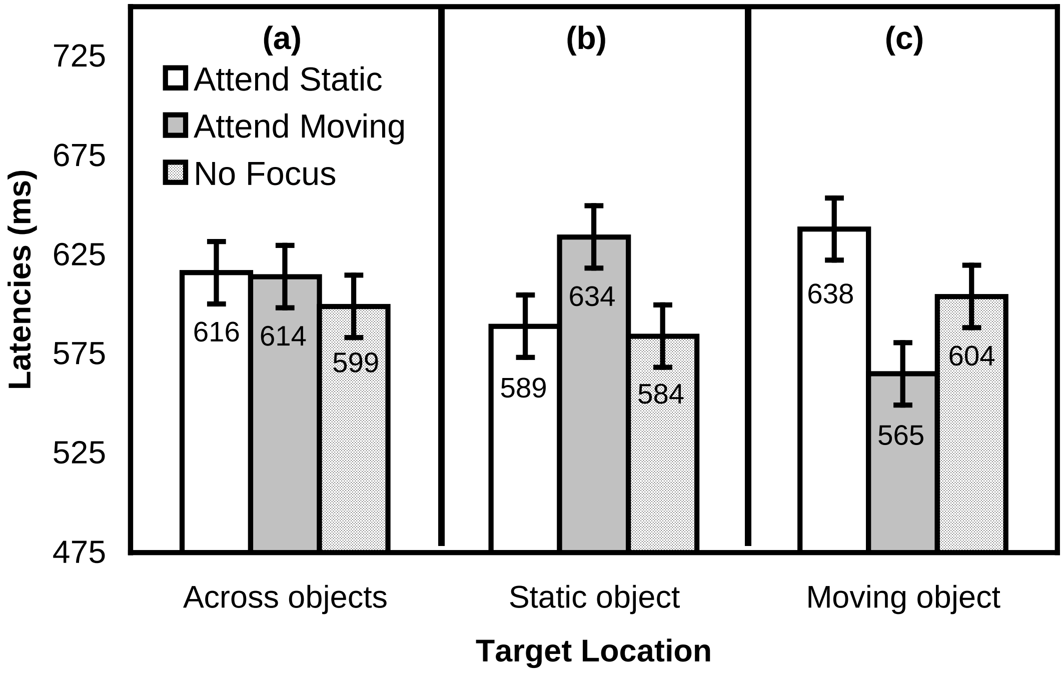

Figure 8: Response latencies (ms) and 95% confidence (15.7 ms) intervals for Experiment 3.

4.5 Experiment 4

Motion is a particularly strong, and developmentally significant determinant of object perception (Spelke et al., 1995), as was noted above; recent studies also suggest that spatiotemporal continuity (a form of common fate) is central to the binding of disparate percepts into a unified object representation (Wallis, 2002), also arguing for the primacy of motion in perceptual organization. Nevertheless, visual stimuli from naturalistic environments, including HUD-equipped cockpits, involve many determinants of perceptual organization, not just grouping by common fate. Also, most work on object-based attention has used a variety of grouping factors other than motion (connectivity, similarity, etc.). The question naturally arises: is motion a more primordial determinant of perceptual processing, and thus of attentional allocation, or is it merely one factor among many? Accordingly, it seemed useful to examine whether adding a grouping factor to moving stimuli would have any “value-added” – i.e., whether, in this case, connecting dots within a group would enhance the object processing advantage over and above the effect of motion. It was decided to investigate this issue by comparing participants’ performance on the 2AFC task (described above) for groups defined by common fate alone with performance for groups defined by common fate and by connectivity as well. To this end, the stimuli used in the preceding experiments were modified so that on some trials the elements within one or both of the groups of dots were connected by lines. (No elements were connected across groups.)

On the assumption that common fate is a privileged grouping factor, one would assume that connecting the dots within groups would have no effect – i.e., the reaction time patterns observed in Experiments 1, 2 and 3 above would be replicated. However, on the assumption that the amount of attention that is allocated to an object depends on the strength or “quality” of the grouping, I hypothesized that there would be an effect of congruence between degree of grouping and attentional focus; that is, connecting dots within a group would enhance the object effect (i.e., the within-object/across objects disparity in performance on the 2AFC task) when that group was the focus of attention, and it would reduce the object effect when the group in question was not the focus of attention. Another way of stating the same hypothesis is that connecting dots within groups would hinder divided attention when an observer spread attention to the display as a whole, and would enhance focused attention when an observer attended a specific object.

Before discussing the results of this experiment, it is worth noting that the experiments up to and including Experiment 4 all relied on the assumption that attentional allocation and perceptual organization is still largely determined by stimulus properties, with top-down influences such as instructions to attend to one object rather than another exerting mainly a modulating effect on stimulus-driven perceptual organization. In other words, attentional allocation and capture are mostly determined by the intrinsic properties of stimuli. As discussed below, the results of Experiment 4 show that stimulus properties alone do not determine attentional allocation in any straightforward way.

4.5.1 Method

Participants

Four participants, with normal visual acuity, colour vision, and contrast sensitivity volunteered to participate in the experiment. All were drawn from the Carleton University community. Two were female and two were male. Ages ranged from the mid-twenties to the mid-forties. A small number of participants was sufficient because, as explained below, each participant provided a large number of data points.

Display and Procedure

The display, task and instructions to the participants were identical to those used in Experiment 3, with the following exception: on some trials, the dots in either the moving group, the static group, or both groups, were interconnected to form a polygon. In the case of the static group, which contained 4 dots, this occasionally resulted in a concave polygon (i.e., with two edges crossing each other) being displayed. There were four connectivity conditions (no groups with connected dots, moving group connected only, static group connected only, both groups connected), which were blocked for each participant (that is, the connectivity condition of the display was set in advance for each block, and was varied between blocks only).

Design and Statistical Analysis

The 4 levels of connectivity, 3 levels of attentional focus (attend static group, attend moving group, no focus) and 3 levels of target location were combined to produce a total of 36 conditions (as target colour, i.e., same or different, failed to produce significant effects in any of the previous experiments, this factor is ignored in the design of the experiment). Target location was varied randomly from trial to trial, with all conditions being equiprobable. Focus and connectivity were blocked for each participant. As the factorial combination of focus and connectivity conditions yields 12 conditions, in order to ensure that each condition was presented a sufficient number of times, the experiment consisted of 49 blocks of 60 trials each, the first block being a practice block and thus excluded from analysis. The connectivity and focus conditions were counterbalanced in order to ensure each participant saw each focus x connectivity combination a total of 4 times. Thus, each cell of the experimental design was presented to each participant for an average of 80 trials, for a total of 2880 data points per participant.

The same statistical methods that were used in the previous three experiments were also used in Experiment 4.

4.5.2 Results

Mean reaction times were computed from the median RTs from each condition for each participant. These were then subjected to a 3 (attentional focus) by 4 (connectivity conditions) by 3 (target location) repeated-measures ANOVA. This analysis yielded only one significant main effect, that of connectivity, $F(3,9) = 6.94$, $MSE = 132.4$, $p = .01$, and one significant two-way interaction, that of focus x layer, $F(4,12) = 15.69$, $MSE = 854.7$, $p < .01$. The two-way interaction is consistent with the results of Experiments 2 and 3, where reaction latencies to the 2AFC task were fastest for targets in the attended object and slowest for both targets in the unattended object and targets across objects.

The means for the connectivity conditions (all other conditions pooled) were: 595 ms for the “connect none” condition, 605 ms for the “connect moving” condition, 595 ms for the “connect static” condition, and 603 ms for the “connect both” condition. Thus, the ANOVA above seems to suggest that, regardless of any other factor, reaction times were significantly slower when just the moving group of dots or both groups had interconnected dots than when no groups or only the static group had interconnected dots. Taking the “connect none” condition as a baseline, this suggests that connecting the dots in the static group alone has no effect, but that connecting the dots in the moving group alone is sufficient to slow down reaction times by an average of 9 ms. To confirm this, data were subjected to a 2 (connectivity for the static group) by 2 (connectivity for the moving group) ANOVA. This analysis produced no significant effects, though the main effect of connectivity for the moving group showed a trend in the expected direction: $F(1,3) = 6.95$, $MSE = 50.4$, $p = .078$. Given the high F value, there is some reason to believe that with more statistical power this result would have achieved significance. More importantly, however, connectivity might have a significant effect on RTs within particular levels of attentional focus.

As the critical hypothesis called for a congruence between attentional focus and connectivity, patterns of reaction times as a function of connectivity were compared within each focus condition using the within-subjects confidence interval method (Loftus & Masson, 1994, 2002). The confidence interval was based on the *MSE* of the three-way interaction from the ANOVA above. Figure 9 shows reaction times as a function of target location for various connectivity conditions for the non-attention condition only.

A visual inspection of Figure 9 shows considerable overlap of the confidence intervals between RTs for different target locations, both within and between connectivity conditions. This shows that: (1) none of the connectivity conditions is sufficient, in and of itself, to induce an object effect when attention is spread to the whole display, and (2) connectivity does not seem to have a significant effect on overall RTs when attention is not focused on an object. However, we know from the 3-way ANOVA above that connectivity does have a significant effect (the main effect of connectivity was significant at the .01 level.) This effect must therefore be due to the effects of connectivity in the other two focus conditions.

Figure 10 shows reaction times for the target locations, within and across connectivity conditions, for both the attend static and attend moving conditions. A visual inspection of the figure reveals that the overall pattern of RTs found in Experiment 3 is preserved. That is, regardless of connectivity, targets within the attended group were always responded to more quickly than targets in the unattended group or across groups. Note, however, that in the ‘attend static’ conditions (Figure 10, top panel), the disparity between attended and unattended targets is smallest for the ‘connected moving’ condition. Also, in the ‘attend moving condition,’ the disparity between attended and unattended targets is smallest for conditions where only one group has interconnected dots, and largest when both groups have the same connectivity condition (both connected or both unconnected). As far as differences for individual target location conditions across connectivity conditions are concerned, note that in the ‘attend static’ conditions, RTs for targets on the static layer follow the general pattern seen for the connectivity conditions above – that is, RTs are fastest when the moving group is not connected, and slowest when the moving group is connected.

The data from Figure 10 are difficult to interpret – the only clear conclusion is that connectivity does not have a major impact on the attentional focus by target location interaction seen on previous experiments. The impact of connectivity itself is not clearly disclosed. To analyze this effect further, Figures 11, 12 and 13 show reaction times for focus conditions across connectivity conditions, analyzed for specific target locations.

Figure 11 shows RTs for targets on the static layer only. In all cases, one notes a general pattern wherein targets on the static layer are responded to faster when that layer is attended, and slower when the moving layer is attended, with respect to the baseline condition where no single group is attended. The inhibitory effect is significant for all connectivity conditions, as in Experiments 2 and 3. Also as in those experiments, the facilitating effects of focusing attention were *not* significant, *except for the case where the static group was interconnected.* Thus, it seems that connecting the static group alone leads to focused attention having a facilitating effect on information processing within the attended static group. It is worth noting also that the ‘static connected’ condition also is the one which leads to the most inhibition of attention to the static group when it is unattended.

Figure 12 plots similar results for the case where the targets appear on the moving group. In this case we see that focusing attention significantly enhances responses to targets within the moving group regardless of connectivity, unlike what was observed for the static group above. However, the pattern of RTs for targets in the moving group is significantly affected by connectivity when the moving group is unattended. First of all, the inhibitory effect (relative to the baseline ‘no focus’ condition) is not significant when no group is interconnected. Secondly, the disparity between RTs targets on the moving layer when the layer is unattended (attend static) and the ‘no focus’ baseline is greatest in the “connect both” condition. The RTs for the unattended targets in the moving group in this last connectivity condition are significantly slower than those for any other condition. In fact, the reaction times for unattended moving targets follow the general pattern for connectivity seen above (fastest when the moving group is not connected).

To summarize: when the targets are on the static group, focusing attention on that group has a facilitating effect only when that group is the only one with interconnected dots. Focusing attention on the moving group always inhibits attention to the static group, but seems to do so the most also when the static group only is interconnected. The general effect of connectivity is significant for targets on the static group only when the group is the focus of attention. When the targets are on the moving group, however, the overall pattern of connectivity is only seen when the other, static, group is attended. That being said, the facilitating effect of focused attention is always significant for the moving group, whereas the inhibitory effect of not attending the moving group is significant for all connectivity conditions except one, when no group has interconnected dots. The effects of connectivity for targets on the static group and targets on the moving group seem to be complementary.

As for reaction times for targets located across groups (Figure 13), we note that focusing attention on the static group significantly slows down responses (relative to the ‘no focus’ condition) in all cases except when the moving group is interconnected. Attending the moving group significantly slows responses relative to the ‘no focus’ baseline only when the static group alone is attended. Note that there appears to be no difference between the ‘attend moving’ and ‘no focus’ conditions in cases where the moving group is connected (‘connect moving’ and ‘connect both’), but that there seems to be a trend towards slowed responses across groups when the moving group is attended when no group has interconnected dots.

In summary, the analyses above suggest that the facilitating effect of focusing attention on a group is not sensitive to connectivity for targets on the moving group, whereas it is sensitive to connectivity for targets on the static group, where the facilitating effect is significant only when the static group only has interconnected elements. The inhibitory effects of focused attention, on the other hand, are relatively insensitive to connectivity for targets on the static group, significantly sensitive to connectivity for targets on the moving group, and sensitive to connectivity for targets located across groups mostly when the moving group is attended. Thus, while the 3 x 4 x 3 ANOVA above did not reveal any interactions between connectivity and focus or target location or all three factors, comparisons using within-subject confidence intervals show that the effects of connectivity depend both on which object is in the focus of attention, and which objects have interconnected elements.

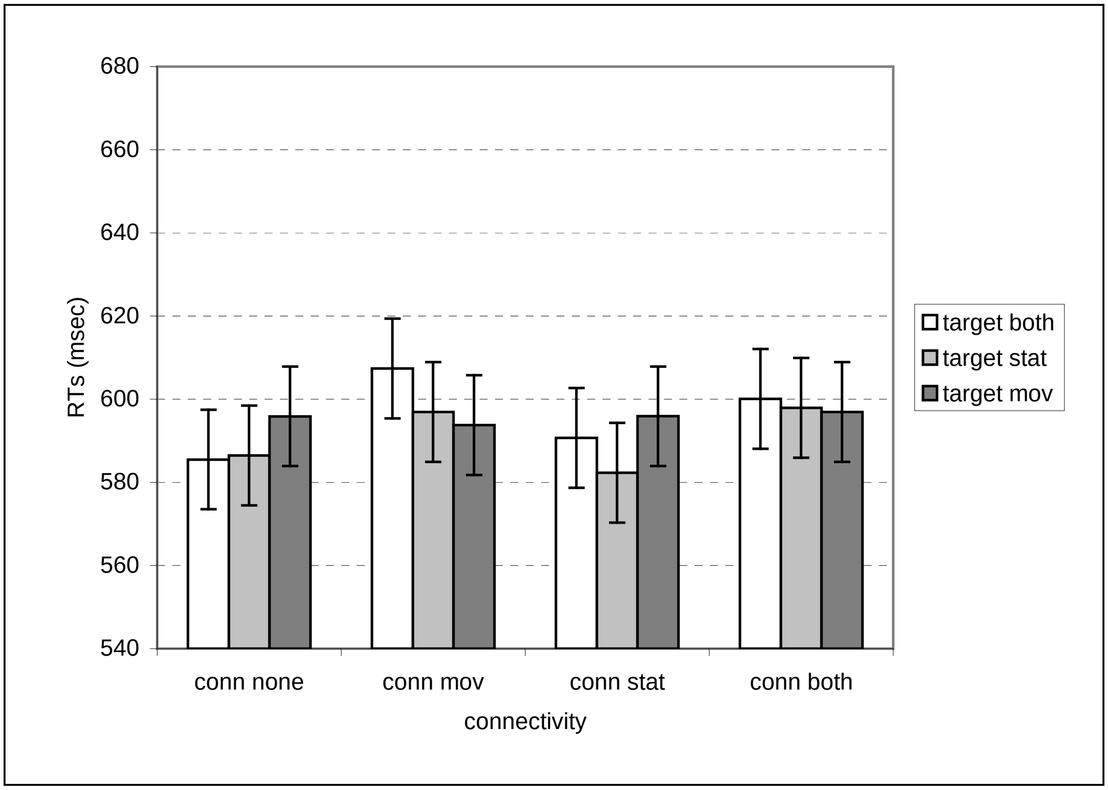

Figure 9: Response latencies (ms) and 95% confidence intervals (11.98 ms) per connectivity condition, Experiment 4, 'no focus' conditions only.

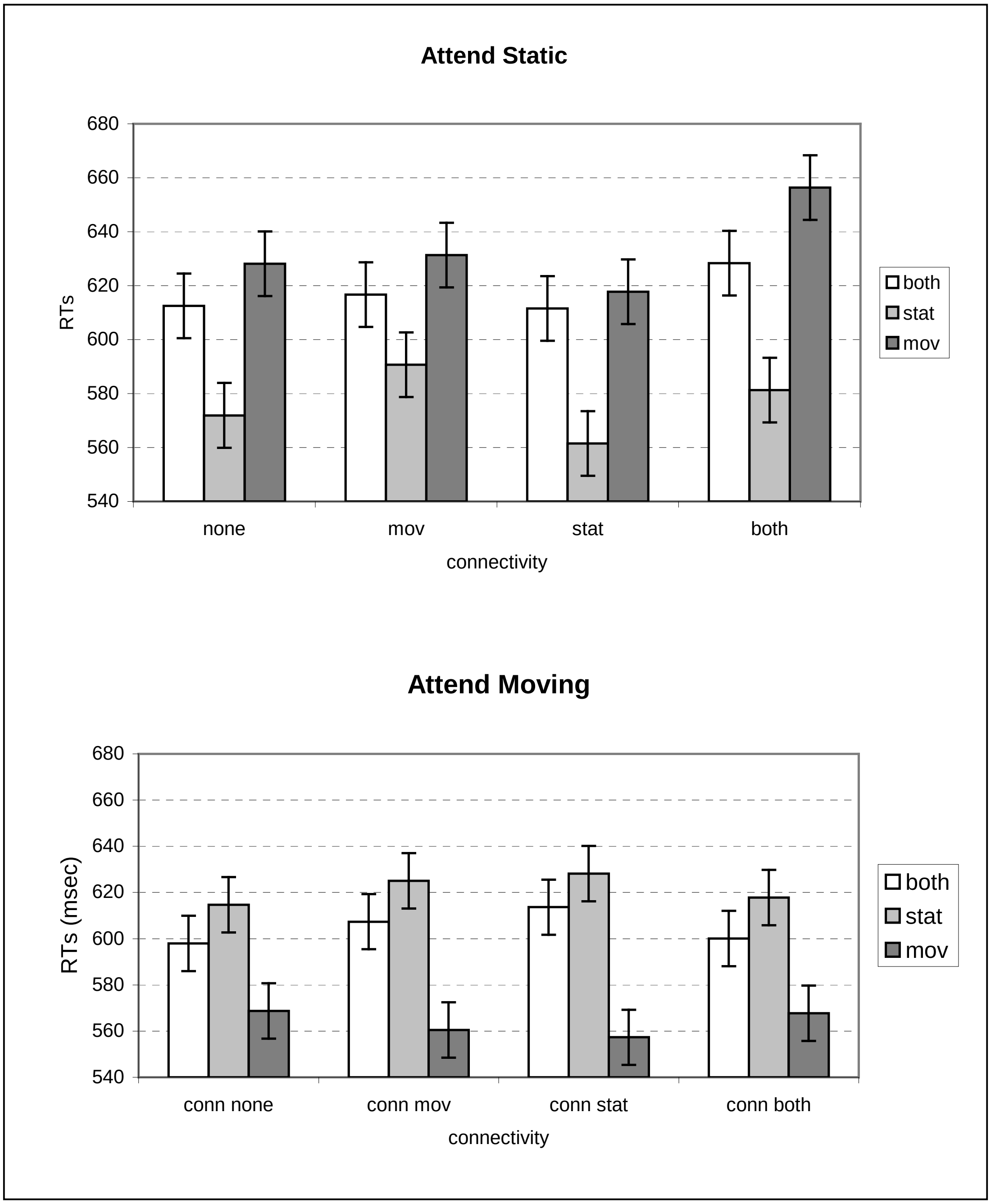

Figure 10: Response latencies (ms) and 95% confidence intervals (11.98 ms) across target locations for different connectivity conditions, Experiment 4. Top panel: Attend static. Bottom panel: Attend moving

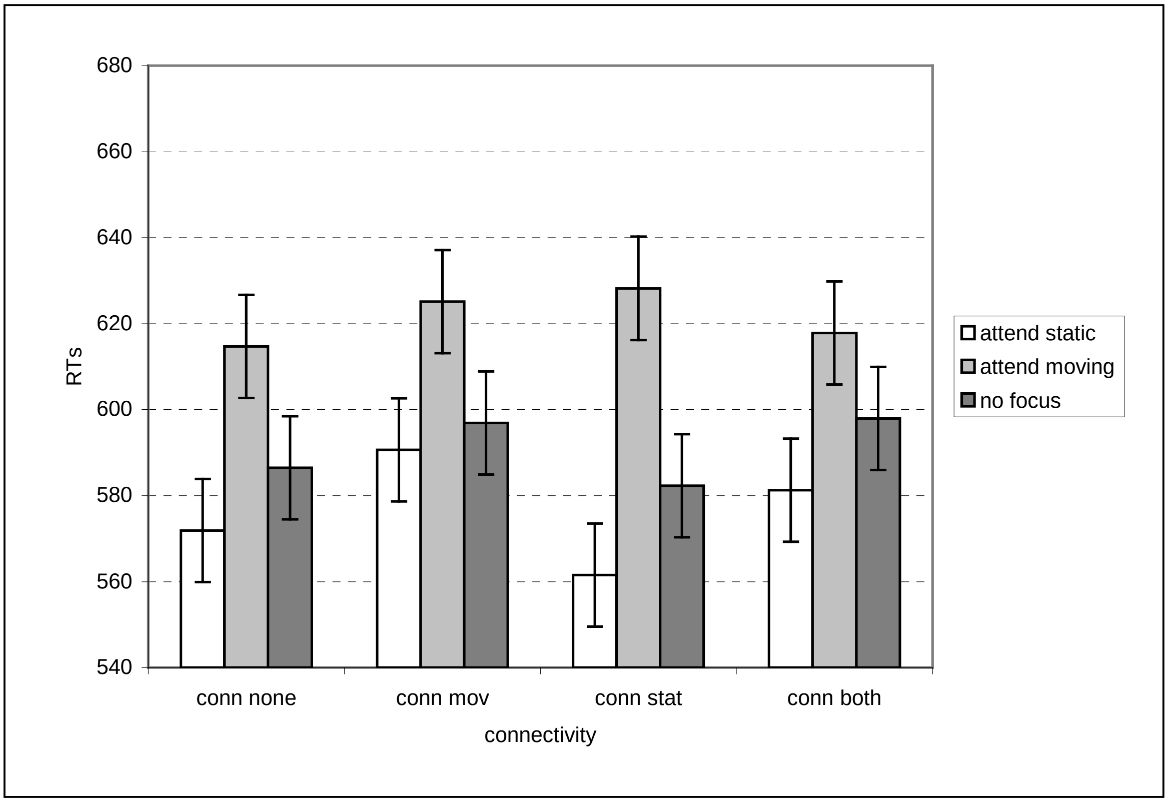

Figure 11: Experiment 4. Response latencies (ms) and 95% confidence intervals (11.98 ms) across focus conditions for various connectivity conditions, targets on static group only.

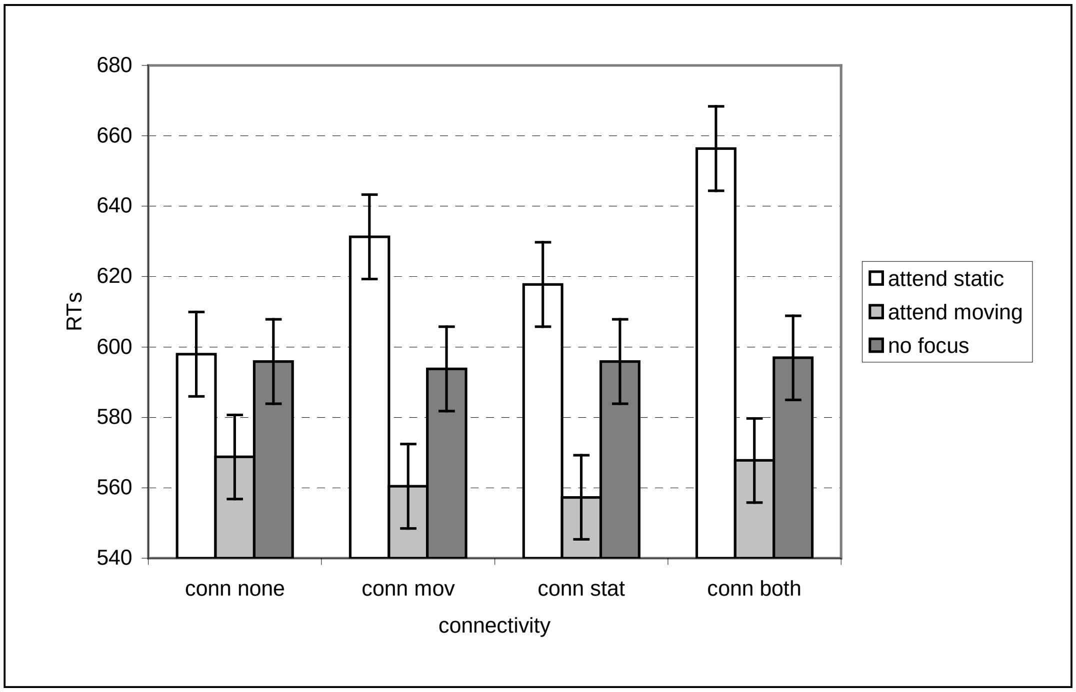

Figure 12: Experiment 4. Response latencies (ms) and 95% confidence intervals (11.98 ms) across focus conditions for various connectivity conditions. Targets on moving group only.

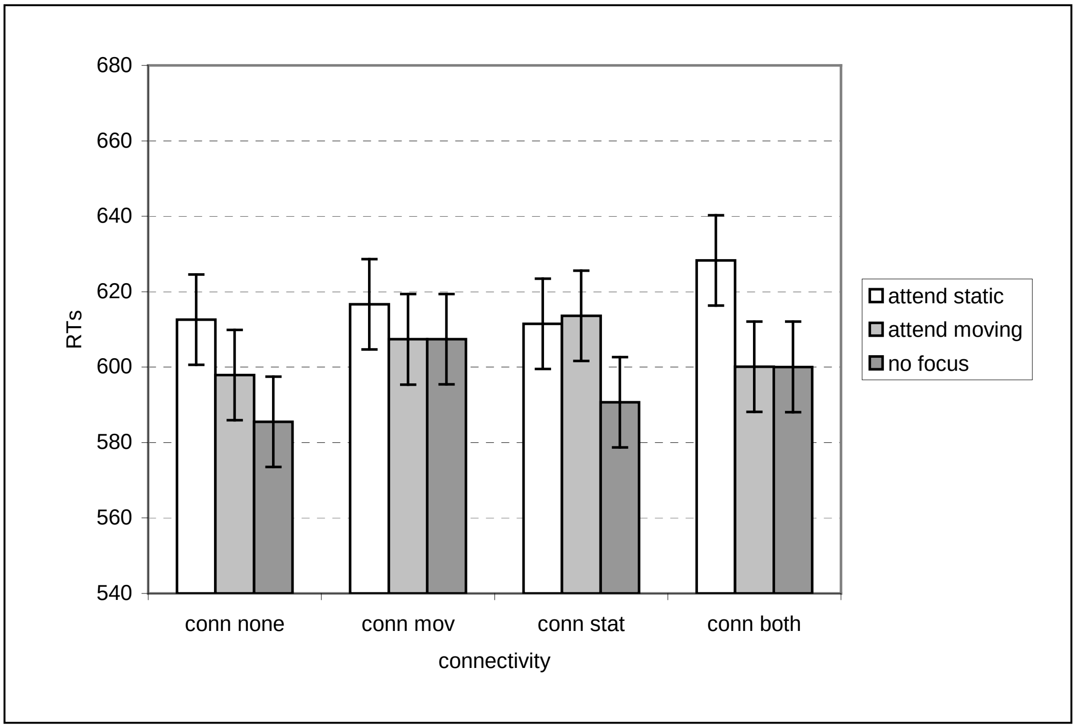

Figure 13: Experiment 4. Reaction times across focus conditions for various connectivity conditions. Targets across groups only.

4.5.3 Discussion

The most obvious conclusions to draw from Experiment 4 is that the focusing of attention onto specific objects generally (perhaps universally) produces significant inhibition of processing of stimuli on objects that are not the focus of attention. In contrast, it enhances the processing of stimuli within the attended object seemingly only under very specific conditions. The results suggest that connectivity can be a factor which, under particular conditions, will enable focused attention to enhance information processing within the attended object, to a small degree. But the conditions under which connectivity can enable the enhancement of information processing are complex and difficult to interpret. Connectivity of elements within groups does not seem to affect the overall effects of focusing attention on a particular group seen in Experiments 2 and 3 in any obvious, straightforward way. However, a closer inspection of the effects of connectivity while holding target locations constant seems to indicate that focused attention is differentially affected by connectivity depending on whether the targets are located on the static or on the moving group. Whereas responses to targets in the static group are enhanced by focused attention only when the static group alone is interconnected, responses to targets in the moving group are enhanced regardless of connectivity. Also, whereas responses to targets on the static groups are inhibited regardless of connectivity when the moving group is attended, responses to targets on the moving group are inhibited by attention to the static group in all but one of the four connectivity conditions (i.e., not when none of the groups are interconnected). When targets are located across groups, attention focused on the static group slows responses regardless of connectivity, but attention focused on the moving group slows responses only when the static – i.e., the unattended group – alone is interconnected.

The results of Experiment 4 are not consistent with the hypothesis that grouping by common fate is a privileged grouping factor that trumps all other grouping factors. However, the results are also not consistent in any straightforward way with the hypothesis that the object effect and the effects of focused attention are amplified by the use of additional grouping factors. Rather, they suggest that connectivity amplifies the effects of object-based attention in a manner that depends very much on which object has interconnected elements and which object is the focus of attention. First, in this respect, is the very notable lack of an effect of connectivity on response patterns across target locations when attention is spread to the display as a whole; that is, when attention is spread out, increased connectivity within groups does not enhance the object effect (note also that grouping by motion also failed to produce significant effects in the 'no focus' condition). Secondly, the effects of connectivity vary depending on which object had interconnected elements: connectivity had an overall effect of slowing responses when the moving group was connected, regardless, seemingly, of whether the elements of the static group were connected. Third, the effects of connectivity modulated the effects of attentional focus in a manner dependent on the group in which the targets appeared. Focusing attention on the static group enhanced processing of targets on the static group only when that group was interconnected, slowed responses to targets on the moving group the most when the moving group was connected, and slowed responses to targets across groups regardless of connectivity. Focusing attention on the moving group enhanced responses to targets in the moving group and inhibited responses to targets in the static group regardless of connectivity, but slowed responses to targets located across groups only when the moving group was not interconnected (the effect was significant only when the static group was interconnected).

In sum, then, connectivity seems to have had the intended effects of (1) enhancing the facilitating effect of focusing attention on an object and (2) increasing the inhibitory effect of focusing attention away from an object *only when the static group is the focus of attention*. It did not seem to have any consistent interactions with object-based attention, distributed or focused, under any other conditions. These results marked a significant turning point in my research program. These results are important in demonstrating that that stimulus-driven perceptual grouping is only one determinant among others of the allocation of attention. Note that the most robust effect in all four experiments has been the effect of focusing attention on a particular object; the effects of motion or grouping in the absence of attentional focus are weak (in terms of magnitude of response latencies) at best. Thus, the remaining experiments described below concentrate on how observers themselves construct their patterns of attentional allocation through their actions and intentions. Stimulus-driven grouping, independently of the observer and his or her context, at most provides very general constraints on attentional allocation, which could possibly serve as default settings for attentional selection in the absence of contributions from an observer’s background knowledge and intentions. Within these constraints, the embodied observer constructs an “attentional structure” through which perceptual information is filtered, integrated and flagged as meaningful.

There are, however, two important limitations of Experiment 4. The first is that connectivity within groups was not manipulated in the absence of other grouping factors. Thus, no baseline for the effects of connectivity alone was established against which the effects of connectivity in other contexts could be evaluated. However, the collected body of evidence on object-based attention using static grouping factors provides good reason to assume that connectivity in the absence of other factors would have produced an object effect. When contrasted with the evidence from Experiment 4, where the effects of connectivity were hard to demonstrate and seemed dependent on target location and focus of attention, this lends credence to the analysis of the Gestalt grouping principles of Palmer (1999) and Pomerantz and Kubovy (1986), who argue that Gestalt principles should at best be seen as “ceteris paribus” laws of perceptual organization – they only explain perceptual organization when no other factors are relevant.

The second, and more serious, shortcoming of Experiment 4 is that of the two groups defined by common fate, one was always moving and the other was always stationary. Clearly this difference had an impact, as the effects of connectivity varied depending on which object was the focus of attention and which object the targets appeared on. One possible explanation for the difference is that attentional selection occurs differently for static and for moving stimuli respectively. McLeod et al. (1989) have argued for the existence of a “motion filter” in the visual system, which would provide an initial parsing of visual stimuli in a scene into all static and all moving elements, respectively. However, this would explain the particular pattern of differences observed between RTs for both groups only if it was also assumed that stimulus-based grouping factors are active only for stimuli outside of the motion filter. While this is a logical possibility, it has not been extensively explored; a study by Cohen (1999) suggests that some global properties of the display might affect whether stimuli in the motion filter are subjected to perceptual grouping, but the exact conditions under which grouping occurs are not fully known. Another explanation is that observers might have organized the display used in Experiment 4 into a figure part (the moving group) and a ground part (the static group). It is known that attentional selection favours figure elements over ground (Vecera, Vogel, & Woodman, 2002). The resulting disparity between responses to elements in the figure and elements in the ground could have then been modulated by connectivity – i.e., if both groups had the same degree of connectivity, the normal figure/ground disparity applies, but if the figure group alone is interconnected, the figure/ground disparity is increased, and if the ground group alone is interconnected, the figure/ground disparity is reduced. While the results above are, I believe, most parsimoniously explained by the figure/ground interpretation, this particular avenue of investigation was not pursued further, for two reasons. First, the role of conative factors – i.e., the observer’s dispositions and motivations – seemed a more interesting and fruitful line of inquiry at the time, and is more relevant to applied settings such as HUDs. Second, the interpretation of the results from Experiment 4 was so difficult that the figure/ground hypothesis emerged only after the following experiments were planned.

Nevertheless, it became clear that, as a stepping stone to studying more fully the conative aspects of attentional allocation, it would first be necessary to investigate how attention is distributed in a display where all objects are defined by common motion – i.e., where groups of elements were distinguished by differing trajectories, not simply by the presence or absence of motion.

4.6 Experiment 5

The ultimate goal of the line of research presented here was to explore the effects of manipulating elements in a display on the organization of attentional deployment in the display. As discussed below for Experiments 6 and 7, an approach was developed involving displays where all elements were moving, but the where motion of some of the elements would be controlled by the observer. Before investigating this factor, it seemed best to first study attentional allocation within displays where all elements were moving, but where groups of elements were distinguished (and presumably differentiated into individual objects) by each having different trajectories. In terms of the standard object-based paradigm, what was being tested was the hypothesis that two moving groups, both defined by a common trajectory, would sustain object-based attention. In the absence of other factors, especially conative factors as discussed in the previous sections, it was expected that grouping by this type of common fate alone would still lead to an advantage in response times for targets within one object relative to targets located across objects. This would establish a baseline against which to examine the role of the locus of control of motion in later experiments.

4.6.1 Method

Participants

Four participants volunteered for the experiment. All had normal or corrected eyesight, and were drawn from the Carleton community. One was male and the other three were female.

Display and procedure

The stimuli used in Experiment 5 were similar to the ones used in previous experiments, except for the following differences: (1) both groups of dots were moving along (distinct) elliptical trajectories; (2) both groups only had 3 dots now instead of one group having 3 and one group having 4 (thus, both groups were distinguishable by the exact shape and direction of their trajectory only); (3) the target dots were displayed in the coloured state for only a brief period (200 ms) rather than remaining coloured until the participant responded. This last change was made both to eliminate possible "higher-order" artefacts (e.g., any perceptual organization occurring after the initial allocation of attention to the display[17]) and to test whether the results found in the previous experiments generalized to displays where the features being judged were only briefly visible. The relative direction of rotation of both groups – either together (iso-rotation) or opposite (contra-rotation) – was also manipulated within participants.

---

[17] It takes on average about 150 to 200 ms to initiate a saccade (Wright & Ward, 1998), yet research such as the experiments of Lavie and Driver (1996) clearly show that object-based effects can occur with regard to stimuli that are displayed within the time it takes to initiate a saccade. Thus, responses to stimuli displayed for longer than 200 ms could well reflect a significant amount of post-attentional processing.

The task employed in this experiment was the same 2AFC task used previously. Participants were instructed not to focus on any particular element or region of the display.

Design

There were 8 conditions in the experiment, resulting from the factorial combination of 2 levels of rotation direction (iso- and contra-rotation), 2 levels of target location (same or different group of dots), and 2 levels of target colour (same or different). All factors were varied within-subjects. Participants were tested over 2 sessions of 7 blocks each (1 practice and 6 experimental blocks), each block having 60 trials. Participants did each session on separate days. Direction of rotation was manipulated across sessions (session 1: iso-rotation; session 2: contra-rotation), so that the iso-rotation data could be analyzed first to verify whether the new stimuli still produced an object effect comparable to the one seen in Experiment 1. The other factors were randomized from trial to trial.

Statistical methods

The statistical methods used were the same as in the previous experiments. One participant did not complete the contra-rotation session, and thus analyses involving the conta-rotation conditions are based on three participants only.

4.6.2 Results

Accuracy rates and correlations

As the display used in this experiment varied importantly from the one used in previous experiments, data were subjected to accuracy and correlation analyses of the type used in Experiment 1. Overall accuracy rates (for 3 participants) were high (over 92%); a 2(rotation) by 2(target location) by 2(target colour) ANOVA revealed no significant effects or interactions. Data were also analyzed for accuracy within the iso-rotation (4 participants) and the contra-rotation (3 participants condition). The 2(location) by 2(colour) ANOVAs for each of these conditions also revealed no significant effects. Correlations of reaction times to target distance at target onset and at participant response, overall and for each factorial combination of the 3 factors of the design, produced low ($r < .1$) and non-significant interactions.

Reaction Times

Mean reaction times were computed from the median reaction times per condition for each participant. The RTs from 3 participants were subjected to a 2(rotation) by 2(location) by 2(colour) repeated-measures ANOVA. The analysis revealed only one significant main effect for target location, $F(1,2) = 18.5$, $MSE = 22.9$, $p = .05$. No other effect achieved significance. However, the estimated observed power for rotation was low (.05), which precludes the rejection of the null hypothesis based on the failure of this factor to achieve significance. As the goal of the experiment was mainly to determine whether an object effect could be found when both objects, not just one, are moving, direction of rotation was not further analyzed.

Since only target location produced a significant effect, data were subjected to a two-tailed paired-samples *t*-test (data pooled across rotation directions and target colour) to confirm the effect of target location found above and to estimate the magnitude of the difference in RTs between locations. The *t*-test was significant, $t(2)=18.5$, $p < .01$, and the mean difference between conditions was 7.97ms, with a 95% confidence interval for the difference ranging from 6.1 to 9.8 ms. Thus, RTs for the within-group condition were on average 7.97 ms faster than RTs for the across-group condition (condition means: 631 ms and 623 ms for across-group and within-group conditions respectively.) The means for each rotation condition are shown in Figure 14.

Since each rotation condition had a different number of participants, data for each of these conditions were also subjected to separate analyses by rotation direction. A two-tailed paired-samples *t*-test of target location for the iso-rotation session (4 participants) revealed a significant effect of location, $t(3)= 4.561$, $p < .05$, and an estimated difference for RTs between target locations of 12 ms (95% CI ranging from 3.6 to 20.4 ms). A similar comparison for the contra-rotation condition (3 participants), also revealed a significant effect of location, $t(2) = 18.5$, $p < .01$, and a difference of 7.97 ms. This analysis suggests that while the effect of target location itself might not have been adversely affected by the small number of participants, the magnitude of the effect itself might have.

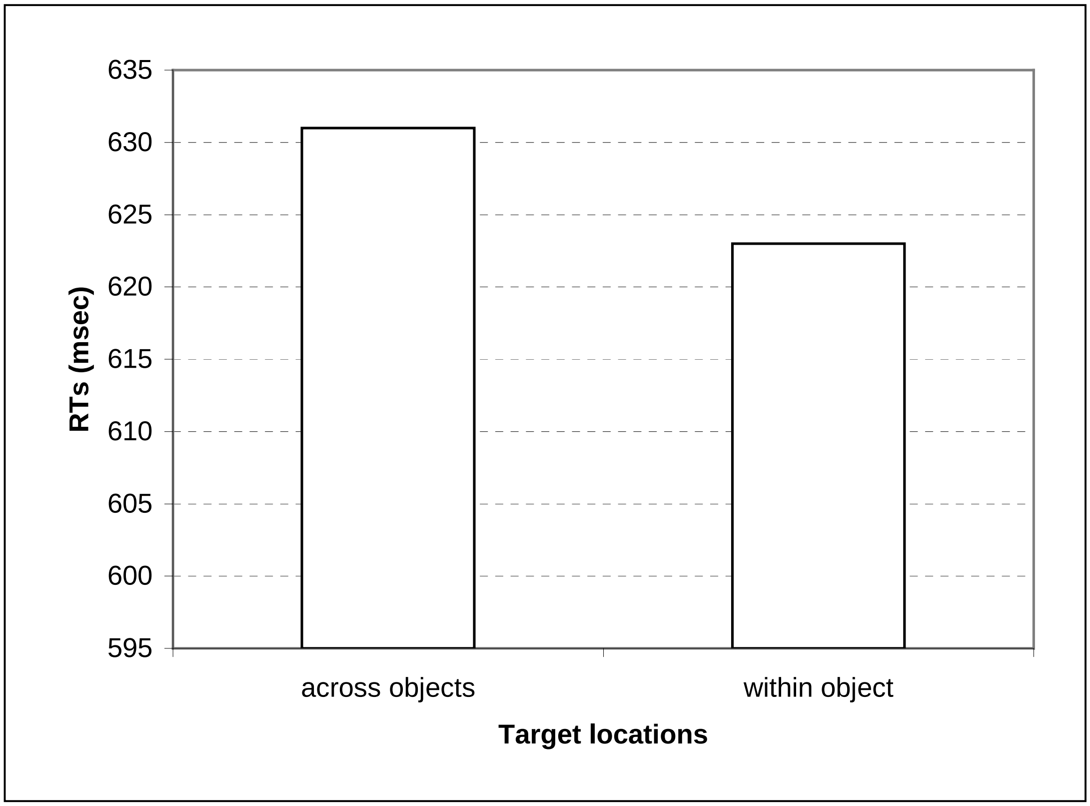

Figure 14: Response times per target location, pooled across rotation conditions and response types, Experiment 5. The difference between location conditions is significant at the .01 level.

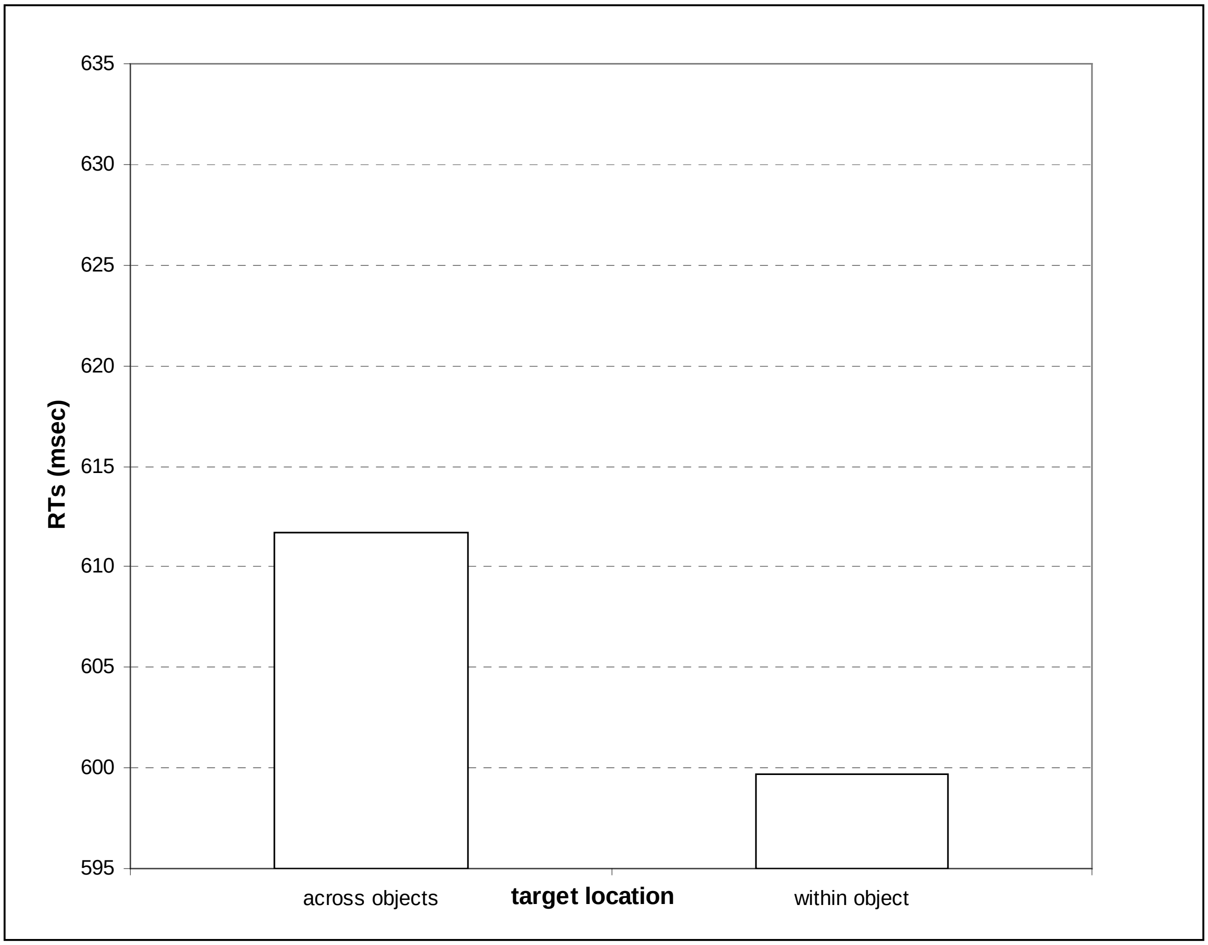

Figure 15: Response times per target location, iso-rotation condition only, Experiment 5, N=4. The difference between target locations is significant at the .01 level.

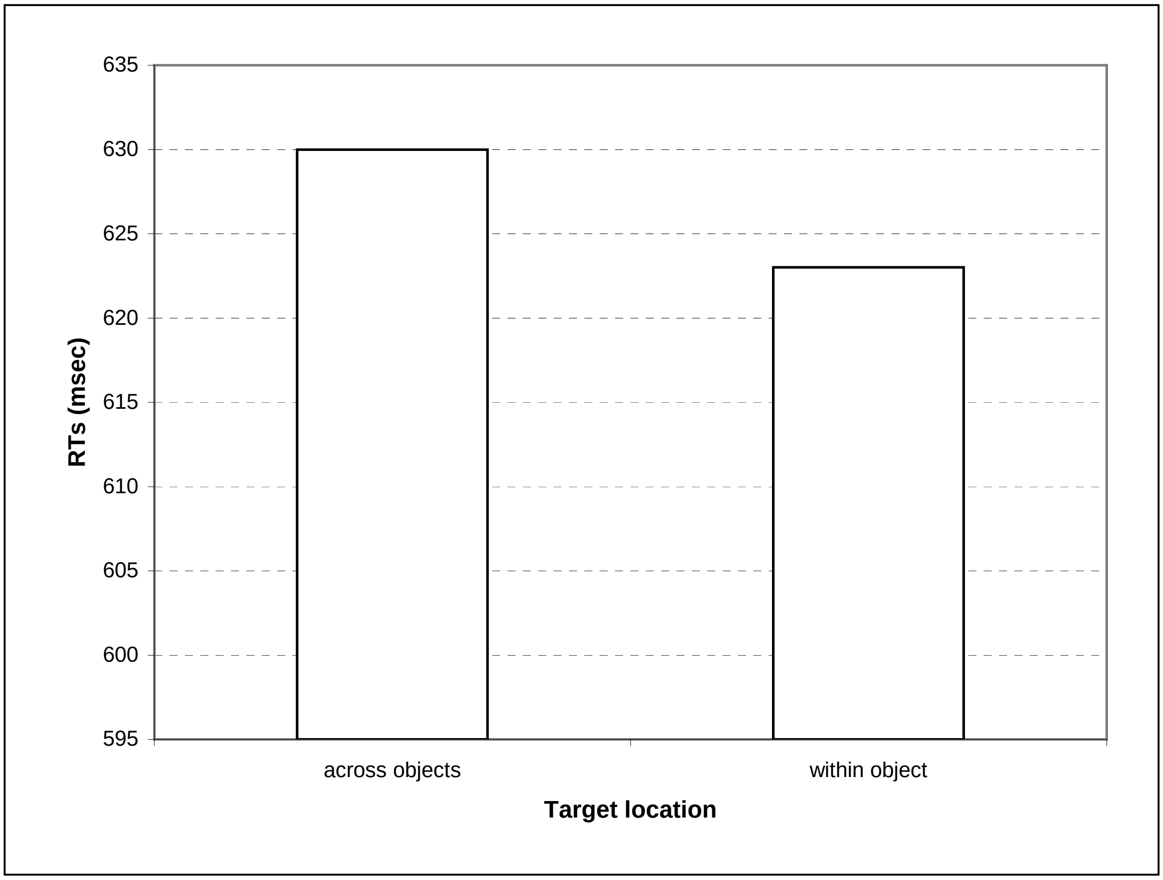

Figure 16: Response times per target location, contra-rotation condition only, N=3, Experiment 5. The difference between target locations is significant at the .01 level.

4.6.3 Discussion

The results of Experiment 4 replicate those of Experiment 1 in its general outline – for a display where two groups of elements are differentiated by distinct trajectories, targets within one group are compared more rapidly than targets across both groups. The magnitude of the effect (8 ms), while small, is roughly consistent with the magnitude found in Experiment 1 (14 ms); furthermore, there is reason to believe that the small sample size precludes a reliable estimate of the magnitude (but not the presence) of the "object effect" seen here.

These results are significant because they replicate those of Experiment 1 with a display that differs from the ones used previously in two important respects: both groups are moving (but differentiated by different trajectory parameters), and the target dots are displayed only for a brief period. This last change suggests that the effects observed in Experiment 1 through 4 do not reflect much "post-attentional" processing. The first change is more significant, because it confirms that the effects seen in previous experiments do not depend heavily on differences in the processing of static and moving stimuli. Furthermore, by showing that object effects can be demonstrated even when all elements are moving, the hypothesis suggested in Experiment 4 whereby moving elements in a display might not be grouped as effectively as static elements due to the "motion filter" (McLeod et al., 1989), becomes less tenable.

Interestingly, the direction of rotation itself seems irrelevant in Experiment 5. On the assumption that the degree of distinguishability of two groups would be affected by their relative direction of rotation, thereby making them more “object-like”, the difference in response times between targets within a group and targets across groups should have been larger for the contra-rotation condition than the iso-rotation condition. The facts that statistical analysis failed to reveal a significant effect of direction of rotation, and that the magnitude of the difference was (non-significantly) *smaller* for the contra-rotation group together suggest that, at least under the conditions that obtained in Experiment 5, relative direction of rotation of the groups has no impact on the distinguishability of the groups and/or the organization of attention in a display. This remains to be tested with more participants, however, as the statistical power for this factor was low.

Finally, the results of Experiment 5, where the motion of both groups is independent of the observer, establish a baseline against which the effects of the locus of control of an object’s movement can be considered. That is, we can determine whether the fact that the motion of some elements in a display is controlled by the user, as in the last two experiments, has an effect on the deployment of attention over and above the simple effect of motion induced by the user’s actions.

4.7 Experiment 6

At the outset of the line of research presented here, I expected that manipulating stimulus-driven perceptual grouping would have a strong effect on the allocation of attention to a visual scene. As Experiments 1 through 5 demonstrate, grouping elements by common motion does lead to a slight processing advantage for stimuli within a group relative to stimuli across groups. However, these experiments also show that more observer-driven factors, such as the intention to focus on a particular group or object, have larger effects on response times than does grouping by motion alone. The only conative (i.e., motivational and volitional) factor studied in these experiments had been, until Experiment 4, the intention to focus attention on a particular group of elements. Experiment 6 was designed to test whether, in the absence of explicit instructions to focus attention on an object, allowing the participant to control the motion of one group would affect the distribution of attention in a display.

The decision to explore the locus of control of motion was motivated by two considerations. First, if a person's actions and intentions to act play a role in directing attention, then the fact that changes in a display are correlated with an observer's actions could make the changing stimuli in question more salient and more relevant to an observer. This could bias attention towards the changing stimuli, and away from other elements in the display. Second, in a HUD, there are many elements that, unlike the environment, respond directly to a pilot's manipulation of his or her instruments, such as altitude and heading gauges, or attitude indicators. Thus, the cognitive tunnelling phenomenon could be partly due to a HUD intrinsically attracting attention to itself because some elements within it are correlated with the pilot's actions. This could lead to an attentional asymmetry, producing stronger inhibition of less-attended elements as seen in Experiments 2, 3 and 4.

4.7.1 Method

Display & procedure

The display used for Experiment 6 was the same used in Experiment 5, with the exception that one of the groups of dots was controlled by the participant via an optical computer mouse.

The computer equipment used for this experiment was a Windows XP-based PC with a 1.79 GHz Intel microprocessor. The computer program controlling the display was written in the C++ programming language, and made use of the Direct-X hardware interface library (Version 8.1) to ensure access to the system clock, video hardware and input/output (I/O) hardware in as direct a manner as possible, bypassing the standard Windows timing & I/O routines, which are not accurate and reliable enough for this type of experiment (Myors, 1999). The input device the participants used was a Logitech optical mouse, with symmetrical left/right buttons, which also ensured timely and accurate motion control and responses. Responses were collected to within 1 ms accuracy. The resolution of the display used was 800 by 600 pixels, displayed on a 17 inch colour monitor, at a viewing distance of 80 cm. Dots in the display were 15 pixels in diameter, and the whole display subtended roughly 12 degrees of visual angle.

Participants were again instructed to perform the same 2AFC task used previously, with the same parameters (colours and duration of display) as used in Experiment 5. Participants were not instructed to focus on any specific part of the display; however, they were given the secondary task of moving one group of dots (using the mouse) so as to keep it at all times as close as possible to the second group, which was moved by the computer program in an elliptical motion as used in previous experiments. Participants’ performance on the secondary tracking task was not recorded, as it was of no theoretical significance; however, to ensure participants actually performed the secondary task, participants were required to start a trial over if during the trial the distance between the centroids of each group exceeded 150 pixels.

Participants and design

Seven members of the Carleton academic community volunteered for the experiment. Five of the participants were female and the other two were male; all had normal or corrected-to-normal vision, had adequate colour vision for the purposes of the experiment, and were right-handed.

Participants were shown 14 blocks of 60 trials each. The first two blocks for each participant were practice blocks, and thus the data from those blocks were discarded. Participants were left completely free to pace their own experimental blocks as they saw fit, and to this end they were given instructions on starting up the experimental software.

Three levels of target location (across both groups, computer-controlled group, participant-controlled group) and two levels of target colour (same or different) were combined to produce six conditions in the experiment. The conditions were varied randomly from trial to trial within participants, so that every factorial combination of location and colour was equiprobable. The trajectory of the computer-controlled group was an ellipse whose direction of rotation and radii were varied randomly from trial to trial.

4.7.2 Results

Accuracy rates and distance correlations

As in most previous experiments, the accuracy rates were consistently high and showed no significant effect of target location or colour. There were no significant correlations between target distance at target onset or at response and response times, thus ruling out spatial effects.

Response Times

Mean response times per condition were computed from the median response times on correct trials per participant. The mean response times were submitted to a 3(target location: across groups, computer-controlled group, participant-controlled group) by 2(target colour: same, different) ANOVA. The ANOVA revealed only a significant effect of target colour, $F(1,6) = 39.9$, $MSE = 281.8$, $p < .01$, with faster RTs for targets of different colour (while significant, this result is of no theoretical importance here). However, a trend towards faster RTs for targets appearing within the computer-controlled group was visible in the data. As the estimated observed power for target location was low (.42, whereas a value of at least .8 is considered adequate, Cohen 1977), the null hypothesis (that there was no effect of target location) could not be rejected, and further analysis of response times per target location was warranted.

Many participants spontaneously noted the task was difficult, as it took some time to get used to using the same input device (the mouse), and thus the same hand, to both control the motion of the participant-controlled group and to click the mouse buttons to make responses. On this basis, data from a further two blocks per participant were discarded; thus, data from the last 10 blocks from each participant were analyzed, for a total of 600 potential data points per participant. Also, one participant exhibited much slower responses than all the other participants (more than one standard deviation from the mean in each condition, see Figure 17). On this basis, the data from this participant was also discarded. The remaining data were again subjected to a 3(location) by 2(colour) ANOVA. Again, only the colour factor achieved significance, $F(1,5) = 29.2$, $MSE = 313.6$, $p < .01$. Target location still failed to achieve significance, $F(2, 10) = 3.16$, $MSE = 174.7$, $p = .086$; however, the estimated power for target location was only .48 (below the .8 required by Cohen, 1977). On this basis, the failure to reject the null hypothesis is possibly an artefact of low power. As the difference between RTs for different colour conditions is not theoretically relevant to this experiment, the data were also pooled for all target colours and submitted to a one-way ANOVA, which again failed to reach significance. The pattern of response times per target location, and the within-subject confidence intervals (Loftus & Masson, 1994) computed with the *MSE* from the one-way ANOVA (94.8) are shown in Figure 18.

An inspection of Figure 18 shows that the responses for the computer-controlled group are significantly faster than responses for targets across groups, as the 95% within-subjects confidence interval overlap by less than half the distance between the interval boundaries and the interval means (Loftus & Masson, 1994.) The difference between responses to targets on the computer-controlled group and the participant-controlled group almost achieves significance by the same method. This analysis also suggests that the failure to find a significant effect of target location in this experiment is indeed a

Type II error. That is, the data of Experiment 6 suggest that target pairs located only within the computer-controlled object are in fact compared more quickly than target pairs located only within the participant-moving group (by about 12 ms) and than target pairs located across objects (by about 13 ms). These findings parallel the results of Experiments 2, 3 and 4, where focusing attention on a single object produced responses to targets pairs within that object that were faster than to target pairs otherwise configured.[18]

---

[18] Note also that the within-subjects confidence interval method of Loftus and Masson (1994) provides a convenient way of making pairwise comparisons between levels of a factor when an ANOVA of the same data fails to reveal significant effects due to low power.

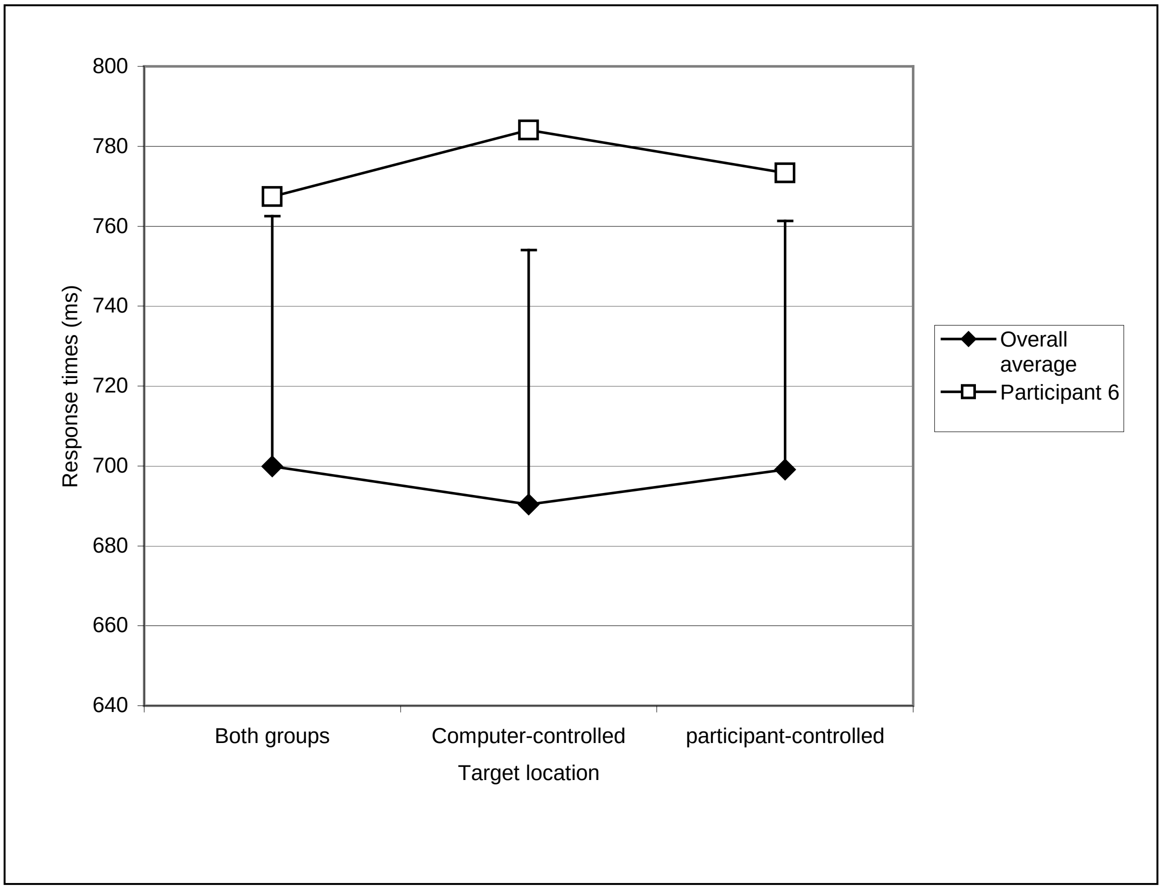

Figure 17: Mean response times per layer condition and response times for participant 6. The error bars represent the standard deviation for each condition.

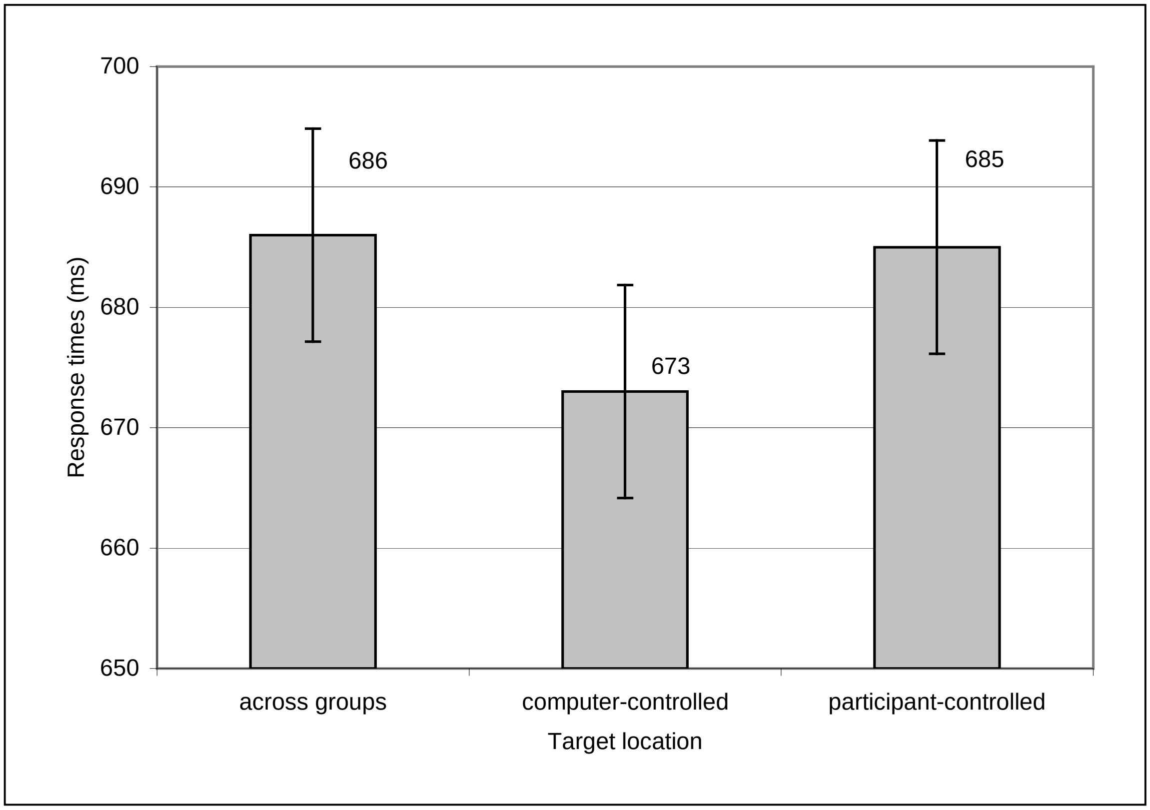

Figure 18: Response times per target location for Experiment 6. The error bars represent within-subject confidence intervals (8.85 ms)

4.7.3 Discussion

Ignoring for a moment the intricacies of the statistical analysis in the previous section, the results of Experiment 6 suggest that participants responded most quickly to target pairs located on the computer-controlled group of dots, and that therefore their attention was asymmetrically divided between the two, with the computer-controlled group receiving the most attention. This result is contrary to the specific hypothesis formulated at the outset of this experiment, viz., that allowing participants to control the motion of one of the group of moving dots would make it more attentionally salient to participants. The results are, however, consonant with the broader hypothesis that the locus of control of motion of elements in a display would alter the symmetrical distribution of attention between two moving groups of dots that was seen in Experiment 5, without eliminating the object effect. Also, the lack of any correlations between target location and response times confirms that the effect is not due to spatial factors, but rather due to some object-mediated attentional organization.

The series of experiments discussed so far does not provide enough information to conclusively state why the computer-controlled group would be responded to fastest. However, a seemingly natural explanation for this effect could be that participants focus on that aspect of the display which is most informative for the execution of their assigned task, i.e., the moment-to-moment location of the computer-controlled group of dots. Thus, these results support the notion that conative factors, in this case task-relevant information, play a decisive role in organizing attention.

The average magnitude of the computer-group advantage (12.5 ms), while smaller than the effect of deliberately focusing attention seen in Experiments 2, 3 and 4, is larger than some of the significant effects seen in the other experiments. There is therefore some question as to the failure to reject the null hypothesis with respect to target location. This might be the result of low power due to a relatively low number of participants (seven). However, the previous experiments reported here had fewer participants than this one (and, with the exception of Experiment 4, had fewer data points per participant). The issue might not simply be low numbers. The design and set-up of the experiment itself likely introduced a large amount of error variance. There are a few factors to consider here. First, the difficulty of the task. Many participants spontaneously reported feeling the task was difficult. On theoretical grounds, this makes sense, as the same response modality (motion of the participant's dominant hand) was used for both tasks. It is likely that there was some motor response interference between the tasks. There might have also been interference from overlearned habits of computer mouse usage (all participants were experienced computer users), although an analysis of error rates per response button, and autocorrelations within sequences of button presses (not reported here) revealed no meaningful patterns. Thus, while the choice of response modalities might not have ultimately obliterated the actual effects on information processing of attentional organization in this experiment, it might well have introduced extra noise in the data. Another consideration is the fact that participants were left completely free to initiate experimental blocks at their discretion, and in some cases participants only completed one or two blocks in a sitting, with many days between sittings. This also likely unnecessarily introduced error variance into the data. Thus, obtaining enough power to an effect of target location that is significant according to a standard repeated-measures analysis of variance would probably involve not just adding more participants, but also re-designing the experiment so as to control for motor response interference (e.g., by using the keyboard as input for the 2AFC responses) and for carry-over effects (fatigue, learning), e.g., by presenting stimuli to participants in two sessions of 6 blocks each (see Experiment 7 below).

However, the data generated with the present set-up should not be dismissed, as an analysis using the within-subject confidence intervals of Loftus and Masson (1994) revealed an effect of target location despite the failure of a standard ANOVA to do the same. Thus, Experiment 6 suggests that manipulating one's visual environment has effects on how one deploys attention. This strengthens the claim that the organization of attention is not driven by stimulus properties alone. Furthermore, the deployment of attention can clearly be determined by factors that are implicit and likely even below a person's threshold of awareness – in other words, the top-down control of attention does not have to involve explicit, deliberate, conscious knowledge and intentions as in Experiments 2, 3 and 4. Both explicit (instruction set) and implicit (i.e., task-driven) strategies operate in the organization of attention. What remains to be determined is how, to what degree, and under what conditions do various contextual and conative factors influence attentional organization. Experiment 6 suggests that informativeness relative to a task will determine how attention is distributed in a display.

As noted above, executing two tasks in Experiment 6 (the 2AFC and the tracking task) might have been difficult due to the fact that both tasks were executed with the same hand. This casts doubt on the experimental design's ability to detect the effects of participant control over the display on attention, a conclusion supported by the low statistical power revealed by the analysis of the results. It might be necessary to eliminate motor code interference between the tasks to fully reveal the effects of self-controlled motion on visual attention. Experiment 7 examined this issue by assigning control for the tracking task to head movements and control for the 2AFC task to the participant's dominant hand.

4.8 Experiment 7

In Experiment 7, participants were instructed to perform the same two tasks as in Experiment 6, but this time they controlled the movement of one of the groups via head movements rather than by using their hands. This had two advantages. Motor interference between the 2AFC task and the tracking task, which might have influenced the results of Experiment 6, was thus eliminated. Also, by displaying stimuli to participants using virtual reality (VR) goggles and by making one group of dots move in unison with the participant's head, the present experiment's set-up simulates the conditions that obtain in a mixed frames-of-reference (FOR) HUD.

Mixed FOR HUDs are HUDs in which a subset of the symbology is yoked to the head movements of the pilot (thus establishing a head-linked frame of reference for some movements in the display), while the rest of the HUD remains fixed relative to the aircraft's axis (thus establishing an aircraft-linked frame of reference). This makes available to pilots some flight information regardless of the orientation of their head, while allowing other information to be visible only when the pilot looks forward (where that information is likely most useful). Such a HUD was designed for use in military helicopters by the Allied Nations' Technical Cooperation Panel 2 (hereafter called the TP2 HUD).

It is difficult to determine *a priori* how the introduction of the TP2 HUD into the cockpit will affect pilot performance, especially given the phenomenon of cognitive tunnelling that occurs with conventional (i.e., fully aircraft-referenced) HUDs. In a study by Herdman et al. (2001a), the TP2 mixed-FOR HUD proved a superior display compared with a standard fixed HUD for certain helicopter training tasks requiring large changes in the heading of the pilot's gaze relative to the helicopter. The TP2 HUD provided pilots constant access to information relevant to their tasks that would have otherwise been available only when pilots were looking straight ahead. This obviated the need for pilots to shift their gaze from the front view and the side view and improved performance on tasks requiring pilots to look to their side for extended periods of time. However, the Herdman et al. study did not address the issue of how attention was allocated within the HUD when information from both the head-referenced and aircraft-referenced parts of the HUD has to be processed simultaneously.

On the standard object-based model, the head-yoked and aircraft-referenced parts of the HUD constitute two perceptual groups, defined by the principle of common fate. Thus, it should be easier to allocate attention to a group in one frame of reference rather than to two groups, each in a different frame of reference. The results of Experiment 6 suggest the additional hypothesis that if participants are given the task of keeping the trajectory of one group of dots as close as possible to the trajectory of another group via their head movements, their attention will be distributed so as to favour processing in the group whose movements they are supposed to follow. In terms of the TP2 HUD, this hypothesis implies that as pilots divide their attention between the head-yoked part of the HUD and other information while making head movements in order to interrogate the aircraft-referenced part of the HUD or the outside scene, their attention will be directed away from the head-yoked HUD symbology. Thus, despite the advantage for the TP2 HUD found by Herdman et al. (2001a), this HUD design might make information presented on the head-referenced part of the HUD harder to process. As this result is initially counterintuitive, and has potentially significant practical implications, it is important to examine whether allowing people to control part of a display always results in less efficient processing of the part of the display they control, as in Experiment 6.

4.8.1 Method

Display and procedure

The display used for Experiment 7 was identical to that of Experiment 6 with the following differences: participants observed stimuli through VR goggles (resolution of 800 by 600 pixels), and they controlled the motion of one of the groups of dots by their head movements as registered by an IS-900 head tracker manufactured by Intersense Inc.

Responses to the 2AFC task were again collected using the same computer mouse as in Experiment 6.

Participants were shown 12 blocks of 60 trials each, over 2 sessions of 6 blocks each. Each session also included a practice block, and lasted approximately an hour. Three of the participants completed both sessions on consecutive days, but two had a lag of 3 days between sessions, and two had a lag of one month between sessions (subsequent analyses showed no effects of amount of time between sessions on performance).

Participants and design

Seven volunteers were again recruited from the Carleton university community. All had normal or corrected-to-normal vision. One participant was female and 6 were male. Only two had significant previous experience with VR goggles. The experiment used a 2(session) by 3(layer) by 2(colour) repeated-measures factorial design. Target location and colour, and the parameters of the trajectory of the computer-controlled group were all randomized from trial to trial.

4.8.2 Results

Accuracy rates and RT/distance correlations

As in previous experiments, there were no significant patterns found for accuracy rates or correlations between target distance and response times.

Response Times

Mean response times were calculated from the median response times for each condition for each participant. They were subjected to a 2(session) by 3(layer) by 2(colour) ANOVA. The ANOVA failed to reveal any significant effects at all (all *F*s < 2.5). An analysis using within-subject confidence intervals, based on the 2-way interaction term from a 2(session) by 3(target location) ANOVA (*MSE* = 197.7), confirms the lack of any significant effects (Figure 14). However, response times per target location do exhibit what looks like a trend away from the expected direction – fastest for the participant-controlled group. Figure 14 also suggests there is a learning effect from session to session – reaction times speed up over time.

A majority of participants spontaneously reported feeling fatigued at the end of both sessions, and a visual inspection of RTs and accuracy rates per session block (Figures 15 and 16) suggests that there was a performance drop-off in the last block of each session. For this reason, data from the last block in each session were discarded. Furthermore, most participants had had no experience with VR goggles and were not used to using head movements to control elements in a display. Therefore, it could be expected that participants would find executing both the head-tracking task and the 2AFC task simultaneously difficult. Thus the data from the first block from each session was also discarded. As the difference in target colour is of no theoretical relevance here, the remaining data were submitted to a 2(session) by 3(target location) repeated-measures ANOVA as above. The ANOVA showed only a significant effect of target location, *F*(2,12) = 5.43, *MSE* = 125.9, *p* < .05. The effect of session showed a trend towards faster overall times in the second session, *F*(1,6) = 4.43, *MSE* = 1223.9, *p* = .08. The relatively low observed power (.42, less than the .8 required by Cohen, 1977) prevents the rejection of the null hypothesis as to the effects of session based on the results of the ANOVA alone. The overall pattern of reaction times, as well as the associated confidence interval (based on the two-way interaction, *MSE* = 137.5), are plotted in Figure 19. The Figure shows that in Session 1, targets on the participant-controlled group are significantly faster than those displayed across groups, and just fail to reach significance relative to the computer-controlled group. In Session 2, both conditions where targets appear within a single group (computer- and participant- controlled) are responded to almost equally as quickly (to within one millisecond), and RTs in both appear faster than RTs to targets across groups. However, all these last differences fail to achieve significance. Note also that, despite the lack of a significant effect for session in the ANOVA, a comparison of confidence intervals shows that, for each target location condition, reaction times were significantly faster in Session 2 than in Session 1.

Since session failed to produce a significant effect, the same data were subjected to a 3(target location) by 2(target colour) ANOVA. This analysis failed to produce any significant effects. Statistical power in this ANOVA was low -- .27 for target location, .25 for target colour, .22 for the interaction between the two. The fact that no effect of target location was found when data were pooled across session, but that an effect of location was found for Session 1 in the analysis above, suggests that the effects of target location did vary with session number.

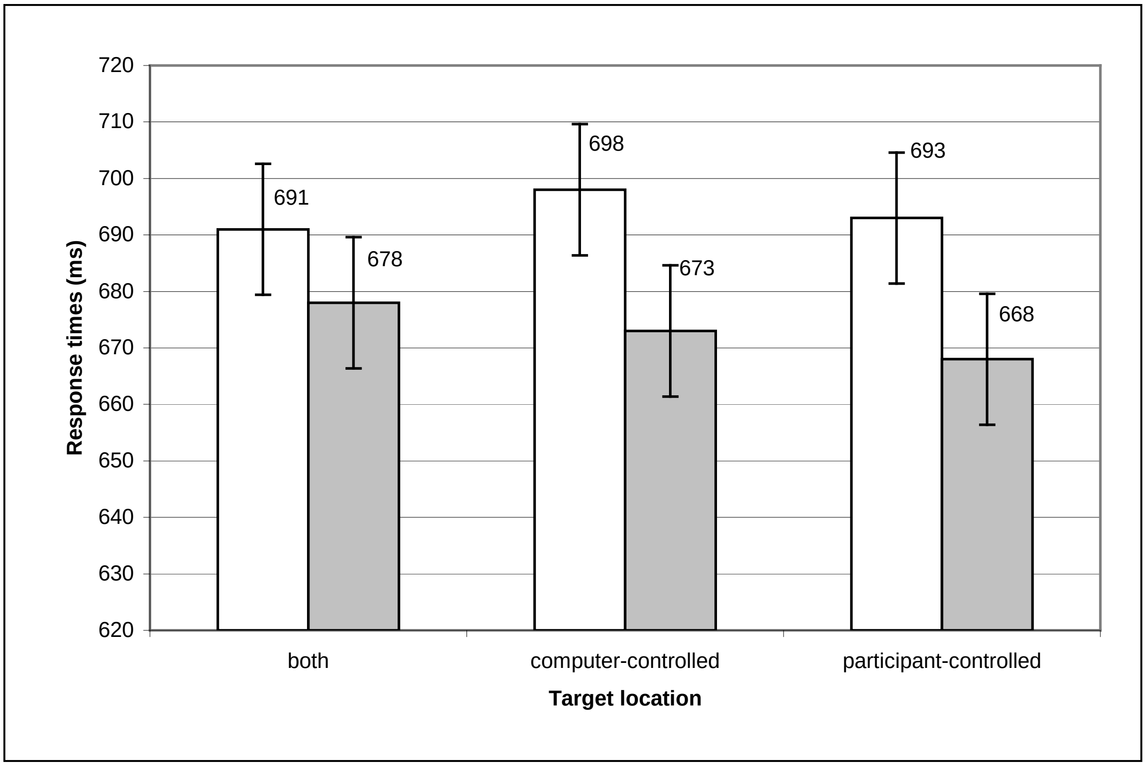

Figure 19: Experiment 7. Reaction times per target location and per session. Error bars represent the within-subject confidence intervals (11.6 ms). Open bars represent Session 1, shaded bars represent Session 2.

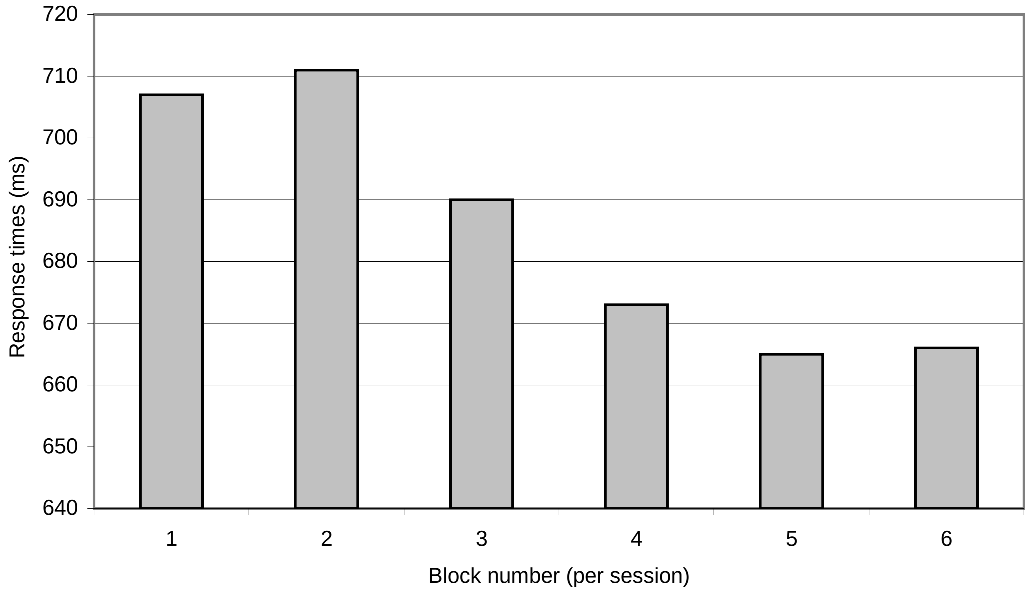

Figure 20: Response times per block, collapsed across sessions, Experiment 7.

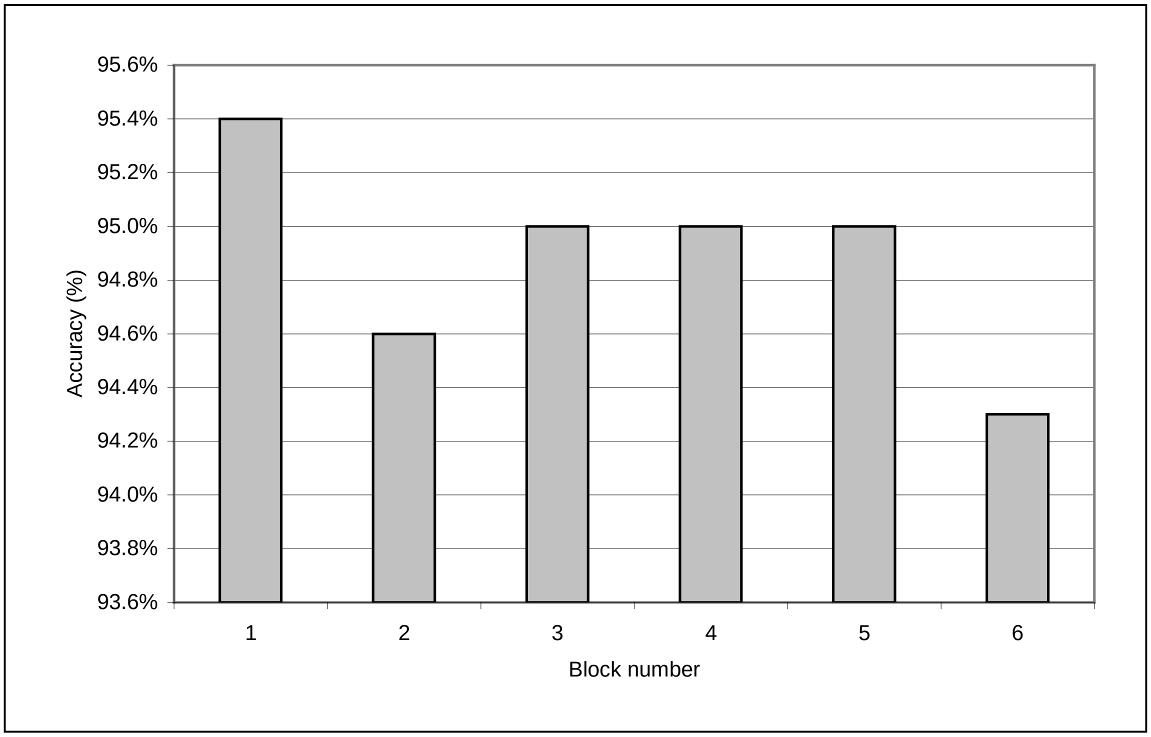

Figure 21: Accuracy per block, collapsed across session, Experiment 7.

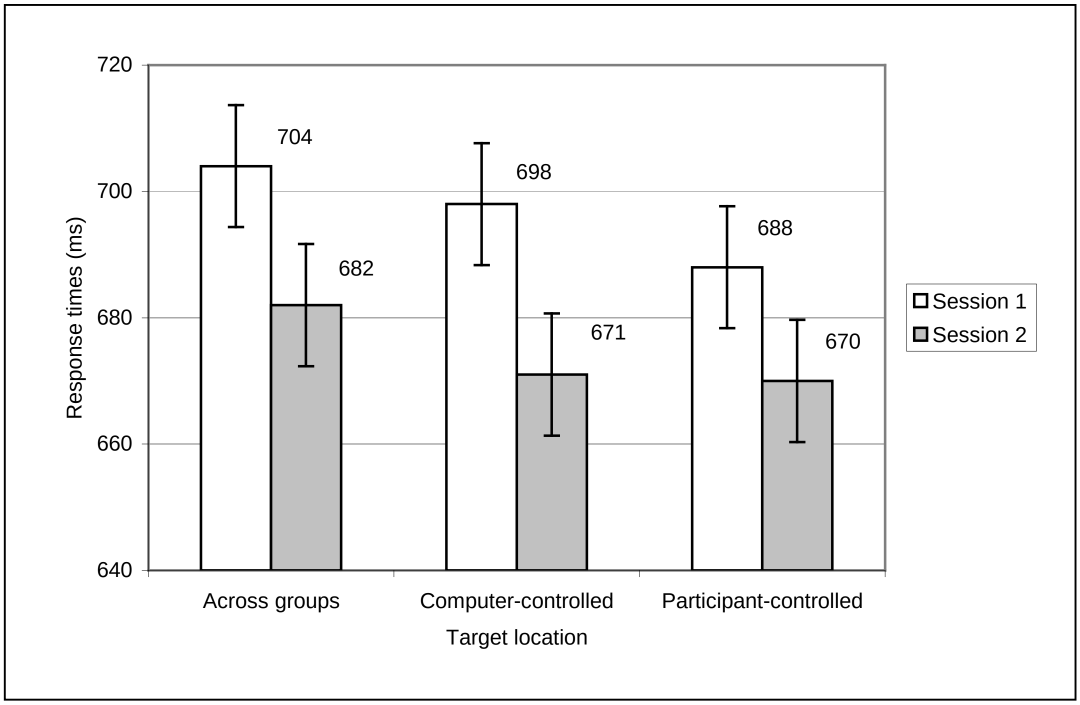

Figure 22: Reaction times per target location for each session, first and last block or each session excluded, Experiment 7. Error bars represent within-subject confidence intervals (9.66 ms). Open bars represent Session 1, shaded bars represent Session 2.

4.8.3 Discussion

The results of Experiment 7 are ambiguous, likely due to low statistical power. The reasons for the low power will be discussed below. It is important to note, however, that despite the difficulty in analyzing the results of Experiment 7, the hypothesis that attention would be distributed asymmetrically in favour of the computer-controlled group was not supported. No main effect of target location was found. While the null hypothesis could not be rejected due to low power, trends in the data were not consistent with the original hypothesis (i.e., targets in the computer-controlled group were not responded to fastest). Furthermore, an analysis of the data by experimental session showed that in the first session, targets on the participant-controlled group produced the fastest responses, and targets displayed across two groups produced the slowest. In the second session, targets within a single group were responded to fastest, regardless of the group they appeared in. The difference between targets within a group and across groups failed to achieve significance, but the pattern is consistent with the standard object effect.

Two broad conclusions can be drawn from the results. The effects of user control over a display on attention vary with the nature of the control, as the pattern of attentional distribution with head-controlled movements differed from the pattern seen in Experiment 6 (manual control). Also, it seems that the effects of user control diminish with time, something that was not seen in Experiment 6.

Why would a head-controlled display produce a different pattern of attentional allocation than a manually controlled display? It is known that the visual system has to compensate for self-movement – that is, movement of the retina due to movement of the eyes, head and body – in order to maintain the perception of a stable world and to accurately track the motion of objects moving in that stable world. As one's head moves in one direction, the retinal image moves in the opposite direction along the retinae. It is thought that recovering the movement of elements in the world therefore involves adding the vector sum of body movements contributing to retinal motion to the motion of elements in the retinal image. It is still unclear how exactly this is accomplished by the visual system, but it is known that vestibular, proprioceptive and visual information all contribute to the process (see review in Harris, 1994). There is also evidence that a copy of the efferent motor signals to the effectors responsible for the self-motion (eye muscles, neck muscles, etc.) – the so-called "efferent copy" – is used by the visual system in this process (Crowell, Banks, Shenoy, & Andersen, 1998). These various information sources are combined to compensate for self-motion in order to recover the motion (or lack thereof) of elements in the environment.

The mechanisms compensating for self-motion in visual perception are optimized for the recovery of the movement and locations of objects that move independently from an observer. As a result, these mechanisms are less successful in perceiving the motion of objects that move in unison with the observer. Any objects already moving in unison with the observer receive an additional motion component from the self-motion cancellation process, the net result being that their motion is overestimated. This is known as the oculogyral effect (Harris, 1994). It is therefore difficult for the visual system to accurately track the motion of head-yoked elements in a display.

The tracking task used in Experiment 7 requires that the motion of two groups of dots be compared and that the discrepancy between the motions be kept to a minimum. However, the oculogyral effect makes this task difficult by introducing error into the visual system's estimate of the motion of the head-referenced group. It is therefore likely that in order for a participant to successfully perform the tracking task, the visual system must allocate more attention to the head-referenced group to compensate for the error due to the oculogyral effect. This is an example of how not only the intention to perform a task but also the manner in which it is performed influence how attention is distributed.

As noted above, the processing advantage enjoyed by information in the head-yoked group diminished and eventually disappeared in the second experimental session. As error rates did not significantly vary across sessions, it is unlikely that the elimination of the advantage for the head-yoked group resulted from poorer performance in the second session. Rather, it is likely that as participants became more adept at compensating for the error introduced by the oculogyral effect, less attention needed to be allocated to the head-yoked group, and thus attentional distribution followed a more typical object-based pattern. This suggests that the distribution of attention within the same display in the context of the same task can vary with a participant's familiarity with the task.

The differences between Experiment 6 and Experiment 7 are important. They show that the allocation of attention must be understood not only in terms of the task in which it is deployed, but also in terms of how the task is carried out. This suggests a complex interplay between perceptual factors, top-down factors such as task goals, and action in the control of attention.

In Experiment 7, participants underwent two testing sessions. Session number was thus a factor in the analysis of variance of the results, which revealed the change in reaction times across sessions. Furthermore, scheduling two sessions with an equal number of trials per session helped ensure comparable rates of learning across participants. This was not the case in Experiment 6, where participants were free to pace themselves. As a result, participants in Experiment 6 took variable numbers of sessions and variable numbers of days to complete the experiment. It is possible that the effects of learning in Experiment 6 were obscured by the self-pacing of participants. Thus the design of Experiment 7 improved on that of Experiment 6 by allowing for the effects of learning to be observed more reliably.

The results of Experiment 7 have potential implications for HUDs, especially mixed-FOR HUDs with head-yoked elements. Contrary to the hypothesis that was initially formulated, attention was not diverted from the head-referenced elements, suggesting that this would not occur in the TP2 HUD either. On the contrary, it is possible that users of the TP2 HUD (or other mixed-FOR HUDs) would unduly focus attention on head-yoked symbology, thus producing a “cognitive tunnelling” effect relative to the other symbology in the HUD, over and above the tunnelling already seen with standard HUDs. However, tunnelling onto head-yoked symbology might fade as the pilot gains experience with the HUD, just as the processing advantage for the head-yoked group disappeared with time in Experiment 7. On a more general level, the findings discussed here again suggest that experience and training have a significant role to play in how attention is deployed in a HUD – symbology design alone will not determine whether a HUD is used efficiently.

On a technical note, the low statistical power of the experimental design raises some issues. The number of participants (seven) was low, but was comparable to the number of participants used in other experiments reported above that revealed significant effects, as well as in other research in object-based attention (e.g., Lavie & Driver, 1996, often used at most 10 participants in their experiments) and in aviation psychology (see, e.g., Wickens & Long, 1995). As with Experiment 6, it is unlikely that low power was simply the result of too few participants. Participants spontaneously reported that the task was difficult and fatiguing. The two testing sessions were long (45 to 60 minutes each). Furthermore, the VR goggles that were used were of entry-level quality, and did not allow for adjustment of interpupillary distance for individual users. Thus the goggles themselves strained participants' visual system. The perception of motion during self-motion also likely made the task difficult, as noted above. There were therefore many factors that could contribute to error variance in the experiment.

In sum, the results of Experiment 7 did not replicate those of Experiment 6, but they did not support a standard object-based attention hypothesis either. A comparison of Experiments 5, 6 and 7 suggests that acting upon a display affects how attention is distributed within the display. However, the contrast between Experiments 6 and 7 shows, in spite of the limitations of Experiment 7's design, that the effects of action on attention depend not only on the goals a participant has in acting upon a display, but also on how the action is carried out and how the display supports those actions. This experiment shows that more experiments examining the effects of action on attention are in order. However, this line of inquiry was not pursued further in the present research, since the original intent of the series of seven experiments presented here had been to study the role of grouping by motion on attention under various conditions.

4.9 Summary and General Discussion

The seven experiments presented above reflect a progression from studying the deployment of attention as a phenomenon determined by Gestalt groupings (the “standard” object-based thesis) to examining how stimulus properties, the intention to focus attention, and task demands combine to dynamically shape the allocation of attention. Experiment 1 established that, in a display with two groups defined by the principle of common fate (i.e. one group of static elements and one moving group), information within a single group is processed more quickly than the same information presented across the two groups, by about 14 ms. Experiment 2 showed that instructing participants to focus attention on one of two objects defined by common fate results in a processing advantage of roughly 90 ms relative to information presented within the group that is not focused on or across groups. Experiment 3 showed that, relative to the case where participants are not favouring one or another group, focusing attention on a single group inhibits attention to information away from the focus of attention (by about 40 ms), and slightly enhances processing of the focused-on information, but only for the moving group (by about 9 ms). Experiment 4 suggests that connecting elements within a group defined by common motion has different effects depending on which group has interconnections and which group is attended. If only the static group is interconnected, focusing on the static group enhances processing within the group; otherwise, the effects of focusing attention are unaffected by interconnectivity for the static group. Interconnecting the moving group slows processing when the moving group is not focused on, but does not enhance processing within the moving group any more than focusing on the group alone. Experiment 5 shows that the single-group processing advantage found in Experiment 1 also obtains when both groups are moving and have distinct trajectories, demonstrating that the effect in Experiment 1 was not likely not the result of a “motion filter” (McLeod et al., 1989). Experiments 6 and 7 show that the allocation of attention can be biased towards one group, at the expense of information presented in the other group or across groups, when the participant is allowed to manipulate one of the groups, much as the focusing of attention does in Experiments 2, 3 and 4. In the case of Experiment 6, manipulating the trajectory of one group in order to keep it superimposed over the other group, using a computer mouse, resulted in a processing advantage of about 10 ms for the group controlled by the computer. In Experiment 7, participants controlled the trajectory of one group with head movements, resulting in a processing advantage of about 12 ms for the participant-controlled group, but in the first half of testing only. In the second half of testing, information within a single group, independently of its locus of control, was processed more rapidly than information across the groups by about 12 ms, but the difference was not statistically significant.

The results of the seven experiments above paint a complex picture of the interaction between the effects of stimulus-driven perceptual organization and top-down factors such as intentions and task demands. In fact, as discussed below, in some cases the relevance of information in the display to the task – neither purely bottom-up nor purely top-down – seems to direct attention. What should be noted first is that no meaningful correlations between spatial separation of targets and response latencies were found. Combined with the fact that in all experiments both groups were superimposed for most of the time, this demonstrates that any single-object processing advantage found above cannot be due to the spatial separation of targets. Thus any spotlight model is ruled out. The other important thing to note is that every single experiment demonstrated some kind of single-object (or, more accurately, a single-group) processing advantage. What varied from experiment to experiment was whether the single-object advantage was symmetrical (did it obtain for both groups in the display, or only for one), and the means by which the asymmetry was induced. Examining which factors induced the asymmetry and which didn't has important implications for the object-based thesis.

In Experiments 1 and 5, nothing in the instruction set or task demands provided a reason for favouring one group over the other. Under those conditions, the single-object advantage was symmetrical, in that when information was presented in only one group, it was responded to more quickly, by about 12 ms, regardless of the group it was shown on. This is consistent with typical results in the object-based attention literature (Lavie & Driver, 1996; Treisman, Kahneman & Burkell, 1983), thus showing that grouping by motion can serve as the basis for the object effect as well as static grouping factors do. In

Experiments 2, 3, 4, 6 and 7, various means were employed which, on the face of it, could have induced an asymmetry of the single-object advantage. These were: instructing participants to focus attention on one group (experiments 2, 3 and 4), interconnecting dots within groups defined by common motion (Experiment 4), and allowing participants to manipulate one of the groups (Experiments 6 and 7). The most robust was focusing of attention. Instructing participants to favour one object over another resulted in responses to unattended target configurations to be inhibited, by anywhere from 50 to 90 ms. The next factor, in descending order of robustness, was the manipulation of a group.

Allowing participants to control the trajectory of one group induced an asymmetry in responses of about 12 ms. The type of control motion – via a mouse, or via a head tracker – also had an effect. In the case where participants used a mouse, the group controlled by the computer – the group they were tracking – was responded to more quickly. In the case where head movements were used, the head-referenced group was responded to most quickly. The most natural explanation is that in each case, the information most relevant to the task was preferentially attended to. The tracking task they performed involved trying to keep one group over the computer-controlled group as closely as possible. “Relevant” information in this case could be the moment-to-moment position of the computer-controlled group, and/or the position of the participant-controlled group relative to the computer-controlled one. When participants used hand movements, the trajectory of the computer-controlled group was most relevant. When head movements were used, however, the mechanisms that compensate for the effects of self-motion on the perception of motion in the world made it difficult for the visual system to reliably track the motion of the head-referenced group, which possibly made this group the most relevant to the task. Even though the results of Experiment 7 were not significant except in the first testing session, they suggest that what counts as relevant to a task can change depending on how the task is performed, not just on what the goals of the task are.

Finally, the effects of interconnecting the elements within a group seemed to have the least effect on the symmetry of the single-object advantage. Overall, interconnecting dots did not make it easier or harder to attend to a single object versus two. Connecting the dots in a group seemed to have an effect only when participants were also instructed to focus on a group. Even so, the effects of interconnectivity were difficult to determine. The most obvious effect was that connecting dots within the moving group seemed to inhibit the processing of information within that group, but only when the group was not attentionally favoured, thus merely increasing the effect of focusing attention. Interconnecting the static group produced enhanced processing of targets – relative to the absence of connections between dots – only on the static group, and only when that group was attended. Otherwise, interconnecting dots seems to have had no notable effects.

What does this all mean for the object-based attention thesis? The object-based model does not address the issue of the symmetry of the single-object advantage. It merely claims that information in one object is processed more efficiently than information presented across objects. Based on this assertion, one should be able to predict, based on a knowledge of the Gestalt grouping principles, which parts of a display group together and will therefore allow information contained within them to be processed better. It is precisely here that we have a problem. Based on the results of the experiments above, we can say that Gestalt principles predict where information will be more easily processed only when no other factors are involved in the visual processing of a display. This confirms Palmer's (1999) claim that Gestalt principles are only *ceteris paribus* rules – they apply only in the absence of other factors. Furthermore, it would seem that, once a grouping is established according to one set of principles, other stimulus-driven factors seem to have little effect on attention (at least as long as the various factors do not conflict). In the case of Experiment 4, interconnecting the dots – thereby, presumably increasing the perceptual cohesion of elements within a group – did not make one group easier to process, over and above the contribution of grouping by motion, in the absence of attentional focus.

Thus, we can say that the default behaviour of attention is object-based – in the absence of other factors, attention favours stimulus-driven perceptual groups. In the presence of other, top-down, factors, this default behaviour is modified. The results above show that focusing attention can favour one group among the many provided by mechanisms of perceptual organization. The object-based attention theory is easily extended to accommodate the effects of focused attention: it is known that spatial attention can deliberately be focused onto a location through instructions (see review in Wright & Ward, 1998), and that focused spatial attention can affect the perception of temporal sequences (Stelmach, Campsall & Herdman, 1997), which is a type of (temporal) perceptual organization. Vecera (2000) has proposed a model where spatial attention is the mechanism by which the visual system selects one among many possible perceptual groupings.

Note, however, that on Vecera's (2000) view, the intention to focus on an object is merely a top-down modulation of bottom-up processes – stimulus-driven perceptual organization provides a menu of objects from which to select, and focused attention then selects from the menu. I propose that the results above imply more interpenetration between perceptual organization and attention. On the augmented object-based model, as per Vecera, once perceptual groups are formed, selecting one through focused attention should have the same effect on processing, regardless of the object selected. Furthermore, on the assumption that focusing attention does not withdraw all attention from the "non-favoured" object, information within the non-favoured object should still be processed more efficiently than information across objects. None of these predictions is upheld by my results. In Experiment 4, the effect of focusing attention clearly interacted with perceptual organization, in the form of connectivity within groups. Yet this interaction was not a function of grouping strength: the two groups did not differ according to strength of grouping, but the interaction of connectivity and focus of attention was different depending on which group was interconnected, and which group was attended. This would suggest that focusing attention not only favours one object among many, but can also affect the way the grouping within the object is processed. As the only relevant difference between groups was overall motion of the group, this might even suggest that focusing attention on an object affects how local and global (i.e., contextual) properties of a display are integrated.

In the cases where one group was favoured, either explicitly (by focusing attention) or implicitly (by task demands), it was found on the whole that target pairs displayed within the non-favoured object were responded to no more quickly than targets displayed across groups. If favouring one group does not completely withdraw attention from the other, and if attention were indeed basted on Gestalt grouping, then even information in the non-favoured group should enjoy a processing advantage over information presented in two groups. It could be argued that the absence of a pair of targets within the attended object cues a switch of attention to the other object – effectively, an instruction to search for targets elsewhere – and that object-based attention should not be expected to account for attentional behaviour during switches or shifts of attention. This explanation still does not account for the lack of a processing advantage for a single, unattended object relative to the two-object case. Whether or not attention switching occurs, in the two-object case, information still has to be integrated across two objects, which on the "standard" object-based view should always be more difficult than selecting information from a single object.

Furthermore, it is unlikely that attention was shifted from one object to the other in these experiments. Shifting attention (not eye movements) in response to a visual instruction, known as symbolic cueing, takes about 300 ms to be maximally effective (Wright & Ward, 1998). In Experiments 1 through 4, targets were present until participants responded, up to a maximum of 3000 ms. In Experiments 5 through 7, targets were displayed for 200 ms only. If "unattended" targets were processed only in response to symbolic cues, then performance on the 2AFC task would have been much worse for unattended targets in Experiments 6 and 7 than in Experiments 2 through 4. This is clearly not the case.

The consequence of the failure of the symbolic cueing hypothesis is that it is unlikely (though not completely impossible) that, in the cases where targets appeared on the non-favoured object, attention was initially entirely on the favoured object and was then shifted. Rather, attention was more likely asymmetrically distributed over both objects (this confirms a study by Johnson & Yantis, 1995, which showed that attention can be distributed in an asymmetrical fashion between two locations at once). In this case, the single-object advantage should still hold for the non-favoured object, though to a lesser degree than the favoured one. As noted above, this is not the case. It is as though stimuli outside the confines of the attended object were not organized into an object. My data do not directly support this conclusion, and further studies with more than one object beyond the focus of attention are required to study this issue.

As noted in Chapter 3, experiments by Mack et al. (1992), and Rock et al. (1992) have been interpreted as meaning that perceptual organization cannot happen without attention. Without going so far, and noting that it is extremely difficult to determine with certainty whether attention is truly absent, I propose that attention does more than just select one object among many possible perceptual groupings, as Vecera (2000) suggests: rather, it shapes perceptual organization itself. Furthermore, on the basis of Experiments 6 and 7, I suggest that the interpenetration of attention and perceptual organization operates in a way that takes into account an observer’s goals and the relevance of stimuli to those goals. The precise manner in which a person’s goals and actions affect attention requires studying attention in a range of tasks and types of control modalities that exceeds the scope of the present research.

The idea that the properties of visual stimuli affect attention in a task-relevant manner is appearing more and more often in experimental research on visual attention. Folk and Remington (1999), for instance, have shown that attention is directed to abrupt luminance changes only when that change is relevant to the task, i.e., the luminance change is due to the appearance of a new object in the display, and the observer happens to be looking for new objects. Craighero, Fadiga, Umiltà and Rizzolati (1999), exploiting the research of Milner and Goodale (1998) on action-based visual perception, showed that displays that are congruent with an intended action are processed more effectively than incongruent displays. Also drawing insight from Milner and Goodale (1998), Cisek and Turgeon (1999) note, based on their research on feature-binding, that attention is a "flexible mechanism for unifying features with other features *based on the context in which they appear*" (emphasis added). Mapelli, Cherubini and Umiltà (2002), showed that a feature comparison task was performed more rapidly for features within an object relative to features not integrated into an object only when "object-related information is *relevant to the task*" (p. 72, emphasis added). Mack and Rock (1998), in their studies on a phenomenon they call *Inattentional Blindness* (IB), note that while many geometric stimuli are not consciously perceived if they are not expected by an observer and their attention is fully engaged in a task involving another stimulus, proper names can be perceived under conditions that otherwise elicit IB, and that participant's given names are almost immune to IB – a case of (personal) relevance directing attention to a stimulus if there ever was one.[19] In a parallel to IB in aviation psychology, Wickens and Long (1995; see also Fadden, Ververs, & Wickens, 2001) note that cognitive tunnelling occurs only when (1) a pilot's attention is engaged by the HUD and (2) an unexpected event occurs in the outside scene. Thus, many researchers have reported the general phenomenon of what could be called relevance-based attention. However, most of these researchers either explicitly profess allegiance to the object-based thesis, or profess no theoretical model of attention at all. As Hacking (1983) notes, experimenters can successfully pursue the experimental study of a noteworthy phenomenon for a long time and still fail to integrate the phenomenon in a systematic theory that has satisfactory explanatory power. In Chapter 5, I sketch just such a theory of relevance-based attention, which I have called a conative model of attention, and I will attempt to show how this theory accounts for experimental findings on object-based attention, spatial attention, IB and relevance-based attention under one unified framework.

---

[19] Mack and Rock's stated conclusion – that stimuli are processed to a deep, even semantic level, before attentional selection is applied, and that this attentional selection results in conscious perception of the stimuli – are at odds with my theoretical claims. However, their results are compatible with my conclusions. In Chapter 5 I will discuss how and why this is so.

5. A NEW MODEL OF ATTENTION: ATTENTION AS ACTIVITY

As previously noted, the original purpose of the current research had been to study the role of the grouping by common fate on the organization of attention, within the framework of the (Gestalt-based) object-based theory of visual attention. The interaction of common fate with three other factors (focused attention, interactivity with the display, connectivity between elements within a group) was examined, with the expectation that they would allow the strength of perceptual grouping to be manipulated, thereby allowing the manipulation of the strength of the “object effect” (the processing advantage enjoyed by information within a single object, relative to information presented across objects). On the Gestalt-driven object-based thesis, the manipulation that was most likely to affect perceptual grouping was connecting dots within a group (stimulus-driven grouping). On the other hand, manipulating attentional focus and the locus of control of elements in the display should have affected perceptual grouping, and the object effect, only if perceptual organization was significantly affected by top-down influences, especially action-based factors.

As discussed in Chapter 4, the results did not reveal a robust effect for the interconnectivity manipulation. However, neither did they reveal that focused attention and manipulating objects determine perceptual grouping independently of grouping by motion. The results are most consistent with the idea that attention influences, and is influenced by, perceptual organization in a way that favours information that is relevant to the actions and intentions of an agent.

Other researchers have also arrived at the conclusion that attention selects or emphasizes information in a relevance-based way. A brief review was given at the end of Section 4.9 above. An important point, also noted above, is that many researchers who have suggested that attention is relevance-based are working from an object-based framework, and some do not explicitly propound any theoretical model of visual attention. The object-based theory does not address the relevance-based aspects of attention at all. There is therefore a need for a theoretical framework that integrates perceptual and relevance-based influences on attention.

In the following sections, a model of attention is proposed that integrates the object-based and relevance-based aspects of attention. It is argued that on this model, attention uses visual "objects" (the products of perceptual organization) as tools to direct and guide action, and that in doing so, attention shapes its "tools".[20] As action depends heavily on an agent's goals, motivations, and needs – themes that have traditionally been grouped under the heading of "conation" in psychology – the model is called a conative model of attention. Thus, it is proposed that visual attention is the interface between perception and conation. A functional architecture for attention that integrates conative, conceptual, and stimulus-driven factors in a single attentional system is also given. Since,

---

[20] The notion of primitive object representations serving as "tools" for the activity of allocating attention is inspired by the notion of "cognitive mediation" by representations of Vygotsky (Wertsch, 1985). Vygotsky's account and its application to attention and cognition in general is discussed more fully in Chapter 7.

on the conative account, perceptual organization and attention interpenetrate each other, an account of perceptual organization that would allow such interpenetration is required. Such an account is provided, which uses inference to the best explanation, not Gestalt principles, as a basis for organizing perceptual groups. As a theoretical test of the conative model, it is argued that the model naturally integrates object-based, spatial and relevance-based attentional phenomena under a single framework. The phenomenon of Inattentional Blindness (IB; Mack & Rock, 1998) is singled out for particular treatment, for, as I argue below, the conative model provides a better explanation of the experimental results that lead to the concept of IB than do Mack and Rock's theoretical suggestions.

5.1 Allocating Attention in Action

As a person engages in a purposeful interaction with his or her environment, some of that person's background knowledge, intentions, and survival needs are called forth. As a result, the person attempts (often successfully) to carry out specific actions designed to satisfy those needs, intentions and knowledge. One aspect of these actions is the visual system's use and organization of visual inputs to direct action. As Milner and Goodale (1998) remark, "vision evolved in the first place, not to provide perception of the world per se, but to provide distal sensory control of the many different movements that organisms make." Within this view, attention serves to intelligently combine just the right bits of external visual stimuli and just the right bits of stored knowledge, with respect to the organism's goals and needs, so as to produce the appropriate motor responses (which, in humans, could also presumably involve “motor responses” of a symbolic nature, such as language, and internal responses such as thought).

In order for attention to both mediate action and be affected by action in the way suggested above, I propose a three-part architecture for attention. Two of the constituents are the two types of determinants of attentional allocation, that is, the stimulus-driven products of early visual perception on the one hand, and the higher-order factors, such as memory and conation, on the other. The third constituent is the mechanism that combines the two types of determinants of attentional allocation into some kind of structure or output that can direct action, which plays the role of Gestalt groupings in the standard object-based attention account and corresponds to the a-objects discussed in Section 3.1. Each is discussed in turn, bearing in mind that the first component (essentially, early vision and perceptual organization) is the one that has been the most extensively researched.

5.1.1 Perceptual Organization in the Service of Attention: It Ain’t Gestalts!

For all the limitations of the Gestalt-group-based approach to visual attention, the “standard” view does get one thing right: attention is constrained in some way by the structure of the environment as picked up by early vision. From this fact follow a number of questions: What are the products of early vision that are involved in attention? How are they formed? And, what relation do they bear to real objects in the world?

Many processes of early vision have been extensively researched, dating back to the landmark studies of Hubel and Wiesel (1962), and it is well known that “early vision," that is, the stages of visual processing immediately following the retina but preceding the concept-laden representations of states of affairs we seem to enjoy as humans, recognizes or extracts luminance contrasts, colours, edges, edge orientations, edge directions of motion, and contours (Hoffman, 1998; Palmer, 1999; Solso, 1994). A number of slightly more complex representations are also thought to be generated, such as surfaces, perceptual groupings, figure/ground separation, Marr's 2 ½ -D sketch (Marr, 1982), and basic 3-D shapes (e.g., geons, in Biederman, 1995). Which of these elements are involved in attention?

The short answer seems to be that attention makes use of surface representations generated by early visual processes. There is indirect evidence for this: many studies have shown attention being allocated to shapes defined by illusory contours, such as the Tipper and Weaver study of IOR with Kanizsa squares reviewed in Chapter 2. Also, a study by Atchley and Kramer (1999) elicited the usual object effect by using targets displayed on the *surfaces* of cylinders rendered to look three-dimensional. Palmer and Rock (1994) have also proposed regions of uniform connectedness, i.e., regions of a visual scene sharing a common feature such as colour or texture, as the entry-level units of perceptual organization.

More direct evidence for surface representations playing the role of a-objects (i.e., representational structures to which attention is directed) comes from studies by Nakayama, He and Shimojo (1995), who, using stereoscopic displays, manipulated both binocular disparity and the co-planarity of targets in an attention-cueing paradigm of the kind originally developed by Posner (1980) to study spatial cueing. The original study by

Posner showed that targets appearing at the cued position (valid-cue condition, representing 80% of the trials) were responded to more quickly than targets away from the cued position (invalid-cue condition). Furthermore, response times increased as distance between the cue and the target increased. Nakayama et al. hypothesized that, as the binocular disparity (i.e., the 3-D distance) between cue and target was increased, participants would respond to targets more slowly. They also hypothesized that this effect of 3-D distance would be reduced if the targets were displayed as appearing on the same 3-D plane, tilted away from the viewer so that binocular disparity between targets was preserved. In other words, if the cue and the target appeared as though they were on a common plane in space, increasing the 3-D separation between them would increase response times less. The results of their study confirmed their hypothesis, leading them to state that “attention … is bound to perceived surfaces” (Nakayama et al., 1995, p. 43). Since surfaces have contours, and since objects are commonly perceived as made up of surfaces (Marr, 1982; Palmer, 1999), the notion that attention makes use of perceived surfaces is consistent with the notion that attention is directed to “objects” in the visual field.

How are perceived surfaces generated by perceptual processes? Researchers understand fairly well how surfaces are built up from simpler factors such as edges, motion, and depth cues. I am quite happy to help myself to the available accounts of surface formation in the visual system. But the crucial point here is that surface formation, and more generally the perceptual processes that interact with attention, are not adequately explained by the Gestalt principles of perceptual organization. The reasons for this were examined in Section 3.1, but because of the prevalence of the Gestalt approach to "objects" in the attention literature, it is important to give a review of the state of the art on perceptual organization. In a later section, I will discuss how the perceptual processes reviewed below are used by visual attention.

Modern vision science got its conceptual foundations with Helmholz's formulation of vision as "unconscious inference" (cited in Palmer, 1999). On this view, the task of vision is to infer objects in the environment and their characteristics from the image projected onto the retina. Furthermore, we generally have no conscious access of the inferential process and the information it uses (Hoffman, 1998; Palmer, 1999). An example of this in everyday life is the perception of whole objects based only on a partial view of them. When one sees only two or three sides of a cardboard box, one automatically perceive it as a whole box, and not as just the three flaps of cardboard that are directly "provided" by the retinal image.

However, the 2-D image available to the visual system is ambiguous, in that a particular 2-D image corresponds to any number of possible 3-D environments. This is known as the "inverse problem" in vision: how does the visual system "reverse" the projection of a 3-D environment onto the 2-D plane of the retinal image so as to "reconstruct" the original (distal) 3-D stimulus, or something like it? For instance, why is it that we will perceive a whole box every time, and not just three flaps of cardboard? The current consensus in vision research is that, due to some combination of innate abilities and experience, we quickly learn to "infer" a complete box from the three flaps. That is, our visual system "explains" the retinal image consisting of three flaps as being due to the existence of a complete box. This approach to vision has become the de facto standard in vision research, both when it comes to studying natural vision systems (see Hoffman, 1998; Palmer, 1999; Pomerantz and Kubovy, 1986) and computer vision[21] (see Kellman & Shipley, 1991; Lowe, 1985; Marr, 1982).

The term "inference" bears some discussion. It is most commonly associated with conscious, deliberate reasoning. As such, it is often subdivided into three general areas: deduction, induction and abduction.[22] In deduction, the soundness of a conclusion always follows strictly from the truth of the premises and the form of the whole argument. The classic example is "All men are mortal; Socrates is a man. Therefore, Socrates is mortal." As long as the premises are true, and they follow certain forms that are known to be valid, the conclusion of a syllogism is guaranteed to be true. In inductive reasoning, a general property is derived in a probabilistic manner from observation of particular instances. An example is a poll: from a small sample of the population, it is estimated with a certain degree of certainty (typically 95%, the "nineteen times out of twenty" we hear of so often) that a given fraction of the population as a whole backs a certain political party, and that another fraction supports another party, and so on. In effect, this type of reasoning says "Based on the subset of the population we've observed, we're

---

[21] There are historically well-known, but empirically unsuccessful alternatives to the unconscioust inference view. The structuralist approach to perception, initiated by Wilhelm Wundt, attempted to introspect the "sensory atoms" that were alleged to make up perceptions, and thereby discover how they were mechanically assembled into objects. A more famous, and not-quite-so-dead, alternative, is the "direct vision" or "ecological optics" approach of J.J. Gibson, who claimed that particular "optic invariants" were available in the inputs to the visual system that completely specified the 3-D structure of the world. Since the unconscious inference view is the dominant approach to vision today, and as the debate between the inferential and ecological views are beyond the scope of this research project, they are discussed no further.

[22] This trichotomy, and the notion of abduction, was introduced into modern logic in Peirce (1878).

95% certain that 30% of the population supports party X; but there's a 5% chance we're wrong." The third type of inference, abduction, was originally formulated by C.S. Peirce as a way of capturing hypothesis generation within the framework of formal logic (Peirce, 1878). However, Peirce's notion of abduction eventually became quite complex and difficult to make sense of (e.g., see Peirce, 1903/1997), and the term itself is now used in different ways in different fields (Thagard, 1988). In Artificial Intelligence, 'abduction' has come to denote the mechanisms by which a cognitive system solves the frame problem – i.e., the problem of deciding which among a set of facts are relevant to the problem at hand (Fodor, 1987; Selman & Levesque 1996). A related interpretation of 'abduction,' also used in AI, is that of hypothesis generation (Peirce's original meaning of the word; see discussion in Thagard & Shelley, 1997). A third meaning of 'abduction,' popularized by Gilbert Harman, is that of inference to the best explanation (IBE; Harman, 1965; Lipton, 1991). On this formulation, abduction involves making inferences guided by the goal of providing the best possible explanation for a phenomenon (Lipton, 1991). As Harman and Lipton have given comprehensive studies of how IBE can account for many aspects of discovery and justification in reasoning, I will henceforth use 'inference to the best explanation' and 'abduction' interchangeably.

Of these three types of deliberate inference, the one that seems best suited as a model for understanding visual inference is abduction. Vision is characterized by many researchers as a system whose purpose is to infer or construct the spatial structures (Hoffman, 1998; Palmer, 1999) or causal histories (Leyton, 1994) which best explain the stimuli presented to the eyes. Thagard (1988, p.53) has also remarked that "as Peirce noticed, and as psychologists … have subsequently confirmed, abduction plays a role even with relatively simple visual phenomena." Furthermore, Lipton (1991) has convincingly argued that IBE is based on contrastive reasoning; that is, IBE is always concerned with explaining why some state of affairs rather than some other obtains. Lipton goes on to point out that the particular contrast to be explained is a function of the explainer's goals and general context. Inference to the best explanations thus a mode of reasoning that is more context- and motivation-sensitive than mechanistic deduction or probabilistic induction. Another way of putting it is that abduction is more relevance-based than deduction or induction, and is thus better suited to the type of inference that the visual system seems to perform.

In any case, accounts of deliberate human reasoning can only serve as analogies for what it is the visual system does when generating surfaces and volumes from edges, motion, depth information and colour. For one thing, it is not clear what would count as "the best explanation" for early visual processes, or even visual perception in general. Nevertheless, vision has been most successfully characterized as some form of inference that applies rules to very basic visual stimuli and infers the 3-D structures that best explain the retinal image. Vision scientists have uncovered a number of such rules and have studied how the rules apply to the elements first extracted by early vision (Hoffman, 1998; Palmer, 1999).

While researchers have described a relatively large number of rules (Hoffman, 1998, lists 35 of them throughout his book and suggests that this is only a fraction of the total), there appears to be an overall pattern to these rules. The majority, if not all, of these rules appear to be either based on regularities in the environment (such as light sources usually being overhead, and the mechanics of rigid and organic motion; see Hoffman, 1998, for a review), or instantiations of what is known as the genericity principle (Albert & Hoffman, 1995; Feldman, 1999).[23]

The genericity principle was proposed by Albert & Hoffman (1995). This principle states that the 3-D structure that best explains the proximal stimulus is one that is as generic as possible; that is, the visual system tends to generate 3-D structures that would produce the 2-D image projected onto the retina with the fewest intervening special circumstances. For instance, a straight line in the retinal image could be the result of the projection of a straight line in 3-D space onto the retina, or it could be the result of the projection of some flat surface – say a disk – being viewed edgewise. In the first case, no change in viewing angle will change the perception that the line is in fact a line. In the second case, a very minor tilt in the angle of the disk would very quickly reveal that what looked like a line was in fact a disk, not a line. The disk produced the perception of a line only under very specific circumstances, whereas the line produces the perception of a line under almost all conditions. That is, it produces the perception of a line generically. Hoffman (1998) gives a number of rules of visual inference that are applications of the genericity principle, such as: “Always interpret a straight line in an image as a straight line in 3D,” or “If the tips of two lines coincide in an image, then always interpret them

---

[23] While there is still debate on whether these “rules of inference” for vision are learned or innate, there is certainly evidence that at least some of them can be shaped or affected by experience (Zemel, Behrman, Mozer, & Bavelier, 2002). I am not in a position here to give a developmental account of perceptual organization.

as coinciding in 3D." The principle of genericity seems consistent with the characterization of the visual system as making inferences to the best explanation, with "best explanation" being replaced by "most generic 3-D structure."

A number of rules of visual inference also seem to encode some knowledge about the basic regularities of our physical environment (Hoffman, 1998). In many cases, these rules can be subsumed under the likelihood principle (Pomerantz & Kubovy, 1986), which is simply the idea that perceptual organization should reflect probable states of affairs in nature. This is similar to the genericity principle, although the two principles can conflict, as Hoffman has shown (1998). Nevertheless, likelihood is often a very good explanation for why certain things might be perceived as being organized in a certain way. For instance, physical objects tend to be relatively coherent and solid masses, and thus when a number of elements move in unison, it is natural that they be perceived as part of the same rigid object. This is the intuition behind the Gestalt principle of common fate. In fact, it has been shown that most, if not all, of the principles of perceptual grouping as described by Gestalt psychologists can often be re-interpreted in terms of the likelihood principle rather than the principle of *Prägnanz,* often with more credible results (Pomerantz & Kubovy, 1986). Also, knowledge about basic regularities in the environment, like the effects of gravity and the position of natural light sources, seems to play a role in visual inference. For instance, Hoffman (1998, p. 30) argues that knowledge about gravity can override the principle of genericity so as to produce the percept of a small block stacked on top of a larger block, rather than a small block unnaturally floating over the larger one. Similarly, the fact that natural light sources are almost invariably above the observer's head leads the visual system to interpret shading in specific ways, to the point of re-organizing an image to make the shading consistent with an elevated source of light. The visual system also seems to make use of information about 3-D structure from depth cues such as texture gradients, perspective, and stereo cues, among others (Hoffman, 1998; Palmer, 1999). Thus, it would seem that visual inference is guided to a large degree by vast stores of (mostly tacit) knowledge about the physical properties of the world.

It appears that not all "visual rules" are equal. Rules about how gravity affects objects can override rules about the genericity of an image. Similarly, motion appears to be able to override static, figural grouping factors (Driver & Baylis, 1989). It is therefore necessary to study how the rules are organized and applied. Specifically, it must be determined whether there is a hierarchy for the rules of vision, whether this hierarchy is flexible or rigid, and how this hierarchy is enforced. Given the current state of research, we still cannot say with certainty how a p-object – a visual representation we interpret as an object in the world – is derived from the rules of visual inference. We do know that lines, edges, stereoscopic depth cues, as well as the grouping factor of common region (Palmer & Rock, 1994), play an important role in the generation of the visual surfaces that constitute the a-objects that are used by the visual system as tools in deploying attention in a scene (Nakayama et al., 1995). However, stimulus-driven perceptual organization cannot fully explain the creation of a-objects, as is discussed below.

5.1.2 Conceptual and Conative Factors in Allocating Attention

It is relatively uncontroversial that concepts and associations stored in long-term memory are involved in visual perception in some capacity; what is more controversial is the degree and the manner in which these factors are involved. Few, if any researchers, however, acknowledge or discuss the role of conative factors – intention, motivation, goals – in visual perception, especially at the level of perceptual organization. Teleological notions such as goals and motivations are notoriously difficult to handle in the computational and reductionistic conceptual frameworks in which most research on vision is carried out (see discussion in Thagard, 1996). Nevertheless, both conceptual and conative factors are essential for the understanding of perceptual organization. In a later section I will argue that the problem of bridging the "concept-free" products of early vision and "higher-order" concepts and conation is solved by seeing attention as an activity.

It would be a mistake to ignore the role of conceptual, top-down knowledge in visual inference. As noted above, the "inverse problem" of vision tells us that the proximal stimulus alone is insufficient for a unique reconstruction of the distal stimulus. Since the bottom-up information provided by the stimulus underconstrains possible 3-D structures, top-down, conceptual knowledge is often required to successfully recover shape and structure from the retinae. We can see this in extreme cases such as the classic picture by R. C. James (Figure 23). In it, most people initially see a number of black splotches on a white background, perhaps representing leaves lying on the ground. Once people are told that the image contains a dog, they tend to notice the dog rather quickly and will generally be able to pick it out every time they see the image again. In other words, the knowledge that there is a dog in the picture allows an observer to construct a unique interpretation of the image that would have been impossible otherwise.

Figure 23: Can you find the dog?

Top-down information is essential for vision in general, not just for the perception of ambiguous displays. If, as it has been suggested, vision is a form of inference to the best explanation, then top-down knowledge is absolutely essential. It has been shown that inference to the best explanation, approached as a deductive computation, is intractable (Selman & Levesque, 1996; Thagard & Verbeurgt, 1998). This is largely due to the fact that the task of automatically generating a set of initial premises to guide the generation of possible explanations, known as a “support set,” is itself computationally intractable. However, if a support set is available for a given problem, explanations can be readily generated, and the best one (according to some pre-determined criterion, such as goodness of fit to data, or number of redundant propositions) can be selected by computational means. Conceptual information could provide a “support set” for vision, in the form of assumptions and heuristic rules about the world, thus making inference to the best explanation a computationally tractable model for vision.

Much work clearly needs to be done on what kinds of top-down information are used by the visual system, and how that information is used. However, even at this early stage a distinction can be made between abstract, context-free information on the one hand, and concrete, context dependent information on the other. The image of the dog in Figure 23 is an example of the latter. An example of the former type of information might be very general perceptual categories, such as surfaces, edges, or even “generic” objects themselves (Feldman, 1999). It might be argued that edges and surfaces are already provided by, or at the very least extracted directly from, the proximal stimulus, as suggested above. However, Feldman (1999) argues that certain primitive assumptions, what he calls “existential axioms,” are required for an inferential approach to perceptual organization to work. Without these, the problem of assigning a unique (or almost unique) grouping to a set of initial elements cannot be adequately constrained, which is simply another form of the “inverse problem.” Thus, it might be the case that the visual system starts with certain innate, very general categories that are used to constrain visual inference, and that these categories are supplemented with others that are learnt through experience as the individual develops. The problem of innate knowledge versus experience notwithstanding, some type of top-down knowledge is needed to constrain the “inverse problem.”

Just as conceptual factors must play a role in perceptual organization (even if it is not clear how that might happen), I argue that conative factors play an essential role as well. It should be noted, first off, that the line between conative and conceptual is sometimes fuzzy. Most of us have had the experience of ducking instinctively out of the way of an object hurtling towards our head, before even knowing what the object was, and feeling very shaken afterwards. Should this case count as conceptual knowledge – i.e., our visual system, and indeed our whole being, knowing from bitter experience that being hit in the head is bad – or as a conative drive – the drive towards self-preservation, in effect our bodies saying “I want to live and be healthy”? That being said, there are determinants of attentional allocation that are clearly conative, i.e., dependent on intentions, goals and motivations. The results of Experiments 4, 6 and 7 discussed in Chapter 4 are an example of this. In Experiment 4, the intention to focus attention on an object (a conative factor) had a much larger effect on the distribution of attention than did the interconnection of elements with each group (stimulus-driven perceptual organization), and modified the effects of interconnecting dots within groups. In Experiments 6 and 7, participants attended more to one or the other of two moving groups of dots depending on which group was more relevant to the task at hand – which would only matter if participants were motivated to execute the task. Other researchers have incidentally found evidence for an influence of conation in attention, while looking for other effects. For instance, Folk and Remington (1999) demonstrated that participants shift attention to the appearance of a new object only if the appearance of a new object is relevant to the task. Other such research was reviewed in Section 4.9.

As with conceptual knowledge, there is the question of how conative factors can influence the products of early vision and attentional mechanisms, supposed to be “encapsulated” from higher-order knowledge (Fodor, 1983). A particular issue with conative factors stems from the every-day observation that there are things in the environment that seem to necessarily and automatically command attention, whereas evidence for such mandatory attentional capture in the laboratory is less conclusive. These findings (or lack thereof) are mirrored by the anecdotal observations of people being so absorbed in one task that the fail to notice other very important stimuli, such as fighter pilots being completely oblivious to warning sirens in the cockpit blaring away at 95 dB during a dogfight. This has led some researchers to the view that there are two general types of attentional mechanisms, bottom-up (stimulus-driven, impenetrable by conceptual and conative factors, and automatic), and top-down (consciously directed by an observer, closely tied to intentions and background knowledge), and that the top-down variety is engaged only in the very specific cases that an observer wishes to focus on a particular object or feature of a scene (Julesz, 1990, 1991).

It is premature to posit two attentional systems, one automatic and the other volitional. I propose instead to distinguish between two types of conative factors within a single attentional system. One type would be the “ethological” factors, i.e., the factors that have been “programmed” into us by evolution for their adaptive value. The other type would be more conscious, concept-mediated conative factors such as conscious intentions and desires. The latter would be active only in the context of specific activities (such as actively monitoring a display for low fuel), whereas the former would always be active to some degree, always on “standby” so to speak, ready to “bootstrap” attentional allocation in the absence of strong conscious conative factors, and would become the dominant organizers of attention under certain stressful conditions. Thus, when a person has specific goals to which they are committed, these conscious conative factors dominate the organization of attention. When such conscious conative factors are weak (e.g., looking at something just for the sake of looking), then the ethological conative factors (orienting towards luminance and colour contrasts, general survival motivations) will serve to bootstrap attention – to give vision some starting point from which to begin organizing a visual scene. When the conscious conative factors become very strong – e.g., when someone becomes obsessed with a goal – the conscious factors will override the ethological factors they are congruent with, while enhancing the effects of the ethological factors with which they are congruent. In fact, the phenomenon of attention becoming focused onto threat stimuli, which have become extra-salient in the context of a stressful or threatening situation, is known as the "reorganization of attention" and has been experimentally documented since the late 1950's (Easterbrook, 1959).

On the account I am proposing, various conative factors would be in competition for control of attention at any given time, and only a few (or one) would eventually win out. I propose that similar considerations would determine which and how many specific "entries" in long-term declarative memory would also participate in directing attention. There are many mechanisms falling under the general labels of 'multiple soft constraint satisfaction' (MSCS; Thagard & Verbeurgt, 1998) or 'winner take all' (WTA) mechanisms, explored both as algorithms for computational vision and as models of neural activity (see Tsotsos et al., 1995, and Vecera, 2000), that could explain how the myriad conceptual and conative factors that could potentially be active at any one time become reduced to a smaller, tractable number of more relevant ones.[24] However, just as conceptual and conative factors play a role in perceptual organization, usually thought of as an early-to-midlevel visual process (Nakayama, He, Shimojo, 1995; Palmer, 1999; Yantis, 1998), stimulus-driven factors play a role in directing the WTA processes towards conceptual and conative factors best suited to the environment and the situation at hand (Vecera, 2000). The following section discusses such mechanisms.

5.1.3 Tying it All Together: A Functional Architecture for the Conative Model

It is generally accepted that "early vision," and thus many aspects of perceptual organization which are supposed to "feed" attention, is a modular system – that is, the processes involved have "hard-wired" specialized knowledge, of the type discussed above, that is both inaccessible by and insensitive to top-down influences such as long-term conceptual memory and conative factors (see Fodor, 1983, for a classic statement in favour of the modularity of perceptual processes). Yet, as I propose, early vision and top-down factors interpenetrate and mutually influence each other to guide the deployment of attention by an organism in a manner that is sensitive both to external stimuli and internal factors. There are two aspects to the proposed interaction of these two types of influences

---

[24] The use of MSCS mechanisms to reduce possible top-down factors to a tractable subset is a variation of the heuristics approach to solving the frame problem (the problem of selecting relevant information among all available information, as described in Fodor, 1987).

on attention. One is the functional aspect, the other is the "implementational" aspect (either neural or computational).[25]

On my account of attention as activity, the functional roles of stimulus-driven (i.e., early vision) and organism-driven (conceptual and conative factors) determinants in the deployment of attention are distinct but complementary. Figure 24 illustrates the various functional roles of the different types of determinants. The stimulus-driven determinants provide a default parsing of the visual field into a-objects, the "tools" that will mediate attentional deployment – most often surface representations as "inferred" by early vision from various low-level visual cues (see discussion above). The organism-driven determinants, on the other hand, provide information that facilitates perceptual organization when stimulus-driven cues are ambiguous (such as the ambiguous image of a dog in Figure 23), and that directs attention to elements or groups of elements that are relevant or congruent with an organism's intentions and needs.

The "quality" cues for organizing visual stimuli into surfaces (something akin to the Gestalt psychologists' notion of *Prägnanz*) will determine the availability of particular groupings as a-objects for attentional selection. That is, the strength with which factors such as the genericity principle, the likelihood principle, and the principle of common region (as seen above) warrant the inference that particular elements belong to the same perceptual structure constrain how easily that surface will be attended to, and

---

[25] Together with the purpose I have proposed for attention – directing visually-guided action onto relevant stimuli – the functional and implementational aspects of visual attention I discuss constitute a three-level analysis of attention of the type that is said to be typical of cognitive science (Dawson, 1998; Marr, 1982).

the degree to which information from that surface will be enhanced by attending to it. Mapelli et al. (2002) have found that the degree to which elements are integrated into a perceptual whole influences the ease with which those elements are compared in a feature comparison task. Also, Experiment 4 in Chapter 4 shows that under some conditions, the interconnectivity of elements within a group of dots influences the degree to which attending to the group enhances the visual processing of that group, as well as the degree to which attending away from a group inhibits information processing in a group. Thus, the various perceptual organization cues discussed above provide a rough initial parsing of a visual scene into elements that the visual system will use to organize attention into an "attentional structure," a hierarchical organization of a-objects, which will guide action, as discussed below.

In the absence of organism-driven factors (or when the competing organism-driven factors cancel each other out, as it were), attention will be guided exclusively by the stimulus-driven cues from which surfaces are inferred, and the stronger the cues for a particular surface, the more attention will be allocated to it, putting the corresponding a-object higher in the attentional hierarchy. This situation would correspond to the "standard" object-based attention account found in the literature. But when organism-driven factors are strong, stimuli most relevant to an organism's needs and goals will receive the most attention. However, the congruence of stimulus-driven and organism-driven factors will determine how easily the visual stimuli relevant to an organism's conceptual and conative context are attended, and how easily irrelevant stimuli are screened out. That is, when organism-driven and stimulus-driven constraints on attention are compatible, visual attention picks out stimuli (objects or features) that are relevant to an organism's intentions and actions in an effortless, automatic-seeming fashion. When the congruence between organism-driven and stimulus-driven constraints is reduced (i.e., perceptual organization does not strongly support the integration of the information relevant to the organism into a single perceptual structure), the deployment of attention is more effortful and less efficient. The effects of the congruence between the various systems of constraints on the deployment of attention is discussed in more detail in Sections 5.2 and 5.3 below.

To complete the account of the conative model of attention, it is useful to examine the mechanisms, both neural and computational, which could underlie the functional interrelation of organism-driven and stimulus-driven determinants of attentional deployment that I have proposed. The discussions of perceptual structure from multiple stimulus-driven cues (Section 5.1.1) and of competing conative cues and congruence of attentional allocation determinants above suggest that whatever mechanisms are involved would have to be able to flexibly resolve conflicts between constraints, some of which are compatible or are in conflict with each other, some of which are harder or softer (i.e., more-or-less easily overridden). Therefore, in very general terms, deployment of attention would have to make extensive use of multiple soft constraint satisfaction (MSCS), that I discussed above as a mechanism for reducing top-down influences on attention to a tractable number. MSCS would have to occur at three levels at least, i.e., at the level of stimulus-driven perceptual organization, at the level of organism-driven determinants, and between the two sets of allocation determinants. These three separate

MSCS processes would operate in parallel, and would be organized in a hierarchical manner, in that the MSCS process for each type of determinant would feed the "central" MSCS process that would ultimately be responsible for the deployment of attention. The central, or bridging, MSCS process would allow initial quick-and-dirty parsings of visual stimuli at the early vision level (i.e., candidate a-objects) to propagate, before they are completely stabilized, to the organism-driven MSCS processes, and would conversely allow conceptual and conative factors, either primed prior to input from early vision or activated by it, to bias and fine-tune the early vision perceptual processes. Thus, as activity in both the stimulus-driven MSCS and the organism-driven MSCS dynamically settles into a structured set of a-objects, the bridging MSCS would allow activity from the other two MSCSs to mutually bias and shape each other. Thus, both stimulus-driven early vision processes and organism-driven conceptual and conative factors constitute autonomous systems, as the thesis of the modularity of cognitive and perceptual processes, which is the generally accepted view in cognitive science, requires (Fodor, 1983; Pylyshyn, 2000). However, these autonomous systems would be highly interpenetrated and would be able to influence each other's activity at very early stages of processing, in ways that are generally ruled out by accounts that suppose that perceptual systems are modular and that early perception is "concept-free."

There are a number of models of attention and/or perceptual organization that incorporate MSCS-like mechanisms, and which therefore lend plausibility to the model I propose. A number of these models are neurally-based, i.e., they are models of actual neural processes and are either supported or inspired by actual neurological or psychophysical research. One such model is the selective tuning model of Tsotsos et al. (1995). This model posits that as competing input stimuli activate neural circuits at higher levels of visual processing, winner-take-all (WTA) processes at these higher levels select higher-level neural representations (be it surfaces, features, shapes, etc.) which send efferent (i.e., feed-backward) signals back towards the input circuits, such that the inputs that activated the winners of the higher-level WTA processes are enhanced and all other inputs are suppressed, thus allowing attentional selection to emerge. This chaining of feedforward and feed-backward signals mediated by a WTA-type process at “higher-order” levels resembles the MSCS-mediated interaction of stimulus-driven and organism-driven factors in my account. However, the selective tuning account leaves the initiation of attentional selection to be bootstrapped purely by visual stimuli, and seems not to make room for conscious volitional and conceptual factors in the WTA mechanisms that eventually result in attentional selection. Also, the selective tuning model seems to result in only one input stimulus or region of the visual field being selected by attention, whereas on my account (and on much available experimental data), attention can be deployed onto more than one object or region of the visual field, more or less symmetrically. The selective tuning model is a essentially a neurally-inspired computational model of how one stimulus gets reflexively selected out of many possible ones based on some unspecified organization of higher cortical processes.

An account of visual attention that bears some similarity to the selective tuning model is Vecera’s (2000) biased competition model of object-based segregation and attention. In this model, processes of perceptual organization as described above produce many “objects” which compete for the focus of attention, again in a WTA-manner; the competition is biased both by spatial attention and by “codes” for familiar objects in long-term memory. On this model, object contours corresponding to the region of space attended are favoured in the WTA process, as are object contours corresponding to the most familiar object codes in long-term memory, and the WTA process dynamically weighs the familiarity and the location biases so as to allow one figure, best satisfying location and familiarity biases simultaneously. This model relegates conative factors to the (potentially volitional) control of spatial attention (which is itself poorly understood) and to whatever motivational and ethological factors might contribute to the familiarity of various objects in memory (and unaccounted for in Vecera’s model). It also assumes that the only conceptual factor involved is familiarity, not accounting for the role that other types of background knowledge or conative factors might also play. The Vecera model and the Tostsos et al. selective tuning model would have a difficult time accounting for how a person would perceive and attend to the Dalmatian in Figure 23, because both models would require that the overall shape of the dog be available for selection as a candidate object before conative and conceptual factors are engaged, whereas information about the contour of the dog’s shape is simply not available without general knowledge of what dogs look like, and/or the intention to find a dog, or at least some kind of organic shape, in the image. On the model I am proposing, such conceptual and conative factors would be available to shape and bias the processes of perceptual organization at the same time as those very processes are attempting to extract global contours from the blobs in the image; also, any partial hints at global contours from the stimulus-driven processes would be able to activate relevant conceptual information about the shape of dogs. Thus, I suggest that my model would handle the Dalmatian in Figure 23, and other complex stimuli, better than the Vecera (2000) or Tsotsos et al. (1995) models.

There are other neural models of attention and perception, or of cortical dynamics in general, which are also consonant to some degree with my model. The Adaptive Resonance models developed by Grossberg and colleagues (Carpenter & Grossberg, 1991; Grossberg, 1994; Grossberg, Mingolla & Ross, 1994) posit resonance-based constraint-satisfaction processes between neural circuits responsible for various aspects of visual processing (see Figure 25 for an earlier, simplified version of the model). As with the models discussed above, the adaptive resonance model is largely a passive one, not explicating how conative factors could actively direct attention, rather than simply allowing attentional selection to emerge out of the mutual adjustment of competing factors. It also leaves unaddressed how conceptual factors influence perceptual organization and attention, focusing more on how the resonant dynamics of stimulus-driven processes account for specific percepts reported by participants in particular psychophysical experiments involving stimulus-driven perceptual grouping. Calvin's account of darwininan cortical dynamics (Calvin, 1996) provides neurological evidence for a generalized MSCS-type mechanism in the human cortex, but does not address specifically how this mechanism would be involved in vision or attention (such is not the stated goal of Calvin's research anyway).

While none of the models reviewed above does justice to the notion of attention as activity that I have developed above, they all provide evidence for the notion that MSCS-type mechanisms are the most likely candidates to implement the functional dynamics of attention I discussed earlier. They also suggest that the conative account of attention I provide could be amenable to verification by computational modeling, as there are many computational techniques (genetic algorithms, neural networks, dynamic systems modeling, Artificial Life techniques) that implement some form of MSCS. One such technique, the neural network approach, has been used to model a variety of cognitive phenomena from scientific discovery to analogical reasoning to legal judgement within the coherence framework of Thagard (1988; Thagard & Verbeurgt, 1998; see also Holyoak & Thagard, 1996). Thus, MSCS seems to be a viable computational and neural mechanism for implementing a variety of cognitive phenomena, including, as I argue, the conative deployment of attention.

In sum, on the conative model, attention selects and emphasizes visual stimuli that are relevant to person's actions and motivations. Multiple soft constraint satisfaction is likely the mechanism that allows this relevance-based selection to happen. I will argue below that this relevance-based attentional system can also account for spatial effects, object-based effects and phenomena such as cognitive tunnelling within a unified framework.

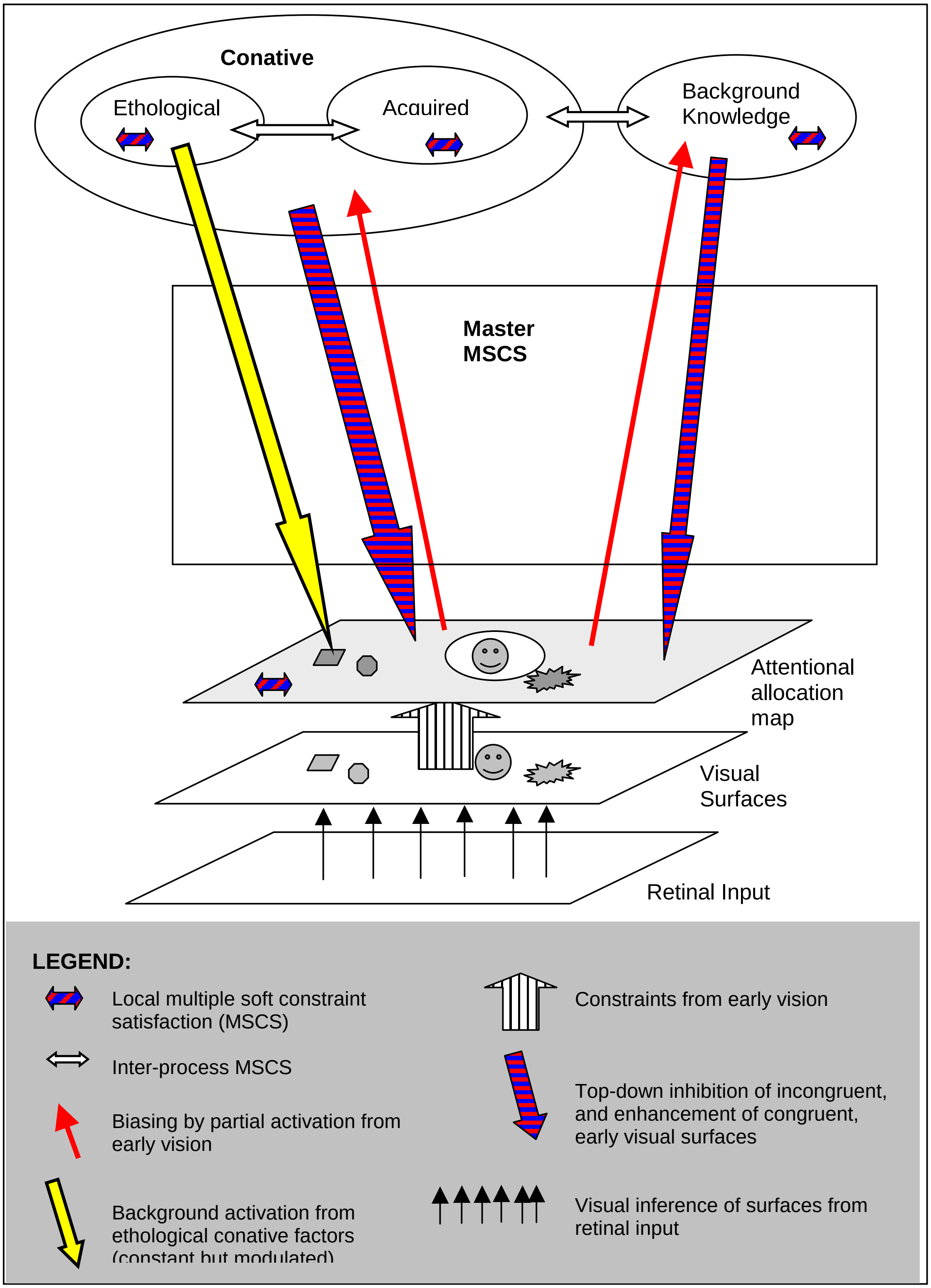

Figure 24: Functional architecture of a conative model of attention

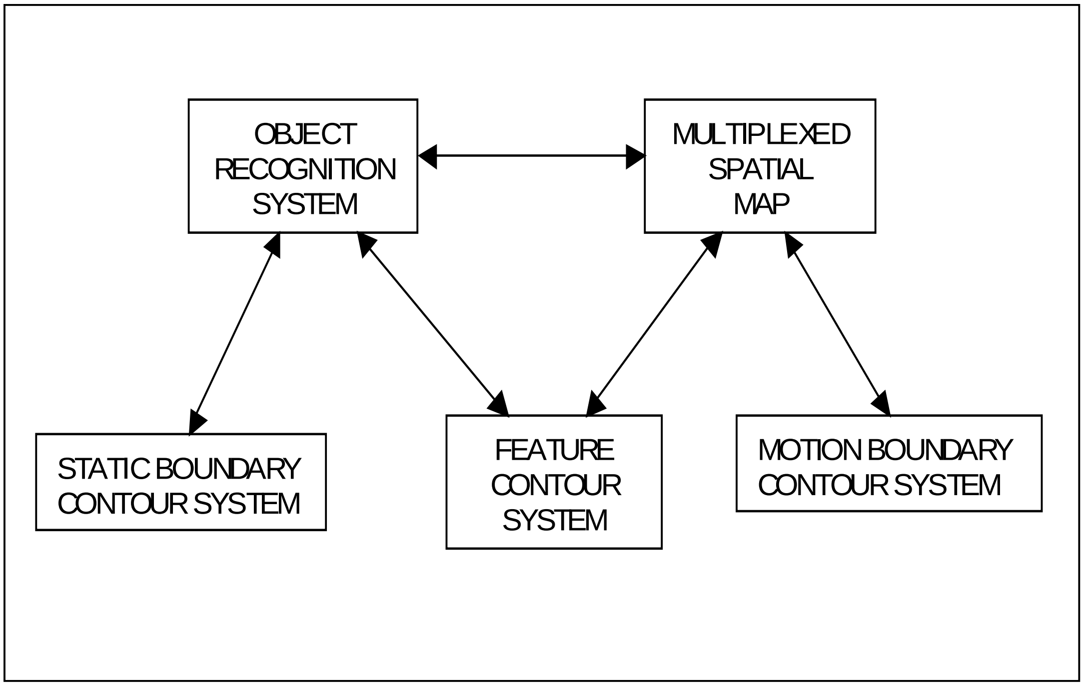

Figure 25: Early version of the Grossberg Adaptive Resonance Model of Visual Perception (adapted from Grossberg et al., 1994)

5.2 Explanatory Power of the Conative Model

It is often said that the true test of a theory is its ability to both explain and predict observed data better than its rivals (Hempel, 1966; Thagard, 1988; Lipton, 1991). In this section I will focus on the ability of the conative model to provide an explanatory framework that integrates spatial attention, object-based attention and relevance-based attentional phenomena.

We saw above (Chapter 2) that current models of visual attention are mostly either spatial or object-based. There have been some attempts to reconcile the two. A few of these subordinate object based-attention to spatial attention. For instance, Lavie & Driver (1996) suggest that attention becomes object-based only once an attentional spotlight has first selected a region of a scene. Logan (1996) reduces perceptual organization to grouping by proximity, which is necessarily a spatial phenomenon, thus supposedly accounting for spatial and object-based attention at the same time. However, other models integrate visual and spatial attention by making spatial attention a mechanism of object-based attention. In Treisman's feature-integration model (Treisman, 1988; Treisman & Gelade, 1980), spatial attention is thought to bind visual features from various "feature maps" into a single percept; on this account, attention is object-based because objects often share features, so attending to more than one object would tend to generate a conflict in the allocation of attention at the level of individual feature maps. As noted above, Vecera (2000) has proposed a model wherein spatial attention serves to bias the winner-take-all competition that selects one object from among the many candidate objects “proposed” by mechanisms of perceptual organization.

While the proposals of Lavie and Driver (1996), Logan (1996), Treisman (1988; Treisman & Gelade, 1980) and Vecera (2000) offer plausible enough accounts of how spatial and object-based attention might be unified, there are two important points that all of these proposals fail to explain. One is the fact that attention seems to play a role in perceptual organization while also depending on it. The other is the relevance-based nature of attention that is being confirmed by a growing body of experiments, as I discussed briefly at the end of Chapter 4. To be fair, with the possible exception of Vecera (2000), none of the accounts listed above was meant to explain these phenomena. However, a comprehensive account of visual attention must address them. Let us examine how these phenomena are explained by a conative model of attention.

5.2.1 The Interpenetration of Perception and Attention

There is evidence that perceptual organization can be affected by attentional factors. Tsal (1999) reports studies that attending to stimuli can alter their perceived length and brightness. Stelmach, Campsall, and Herdman (1997) report that focusing attention alters the perception of the temporal order of events in a display. Mack et al. (1992), Rock et al. (1992) and Ben-Av et al. (1992) report studies where perceptual groupings were not detected when attention was withdrawn from them. Egeth and Moore (1997), arguing against Mack et al. and Rock et al., performed experiments where the Ponzo illusion (also known as the railroad tracks illusion) and the Müller-Lyer illusion could only be perceived if perceptual grouping by similarity of elements in the display occurred. Egeth and More reported that participants were able to perform a task dependent on perceiving the illusions even when they did not attend the grouping patterns and had no awareness of them. However, participants' accuracy in a task involving the Ponzo illusion dropped to chance levels when they were made aware of the presence of the illusion. Thus, attending to the grouping patterns altered the perceptual organization of the display for those participants. Finally, as discussed in Chapter 4 on my experimental results, interconnecting the dots within perceptual groups defined by common motion had no effect on the allocation of attention when participants did not attentionally favour a particular group, yet the same manipulation either enhanced or inhibited attention to the group, depending on whether participants focused attention on that group or away from it (Experiment 4).

On the standard object-based model, these results should not obtain. Since attention is thought to operate at a stage following perceptual organization, attention should not be able to alter perceptual organization, but only select its products. On this view, perceptual organization is "attentionally encapsulated." On Vecera's model (2000), where spatial attention biases the WTA mechanisms of perceptual organization, it could be argued that attention can modify the perceived brightness and length of objects, as per Tsal's results (1999). However, this would not account for the more pervasive effects of attention on perceptual organization that the results of Mack, Tang, Tuma, Kahn, & Rock (1992), Rock, Linnett, Grant, & Mack (1992), Egeth and Moore (1997) and Experiment 4 (Chapter 4) have revealed. The effects on temporal perceptual organization reported by

Stelmach, Campsall and Herdman (1997) would not do much to salvage Vecera's account, as it is not concerned with the perception of temporal order.

On the conative model, attention and perceptual organization are intertwined, as was discussed in Section 5.1.3. It could therefore be said that on this model, attention is object-based for the simple reason that it is involved in perceptual organization. Just how attention might affect perceptual organization under specific conditions is discussed below (Section 5.3).

It should also be noted that the cycles of progressive re-entrant processing between perceptual organization and attention would allow any factors that affect the operation of attention (such as conceptual or conative factors) to thereby affect perceptual organization. This is the basis of the relevance-based effects seen in research on attention, as I discuss next.

5.2.2 Relevance-Based Attention

The idea that cognitive processes exploit the relevance of information and stimuli to cognitive agents is gaining acceptance in cognitive science. An example of this relevance-based approach to cognition that is generating much excitement is Relevance Theory (Sperber & Wilson, 1995). This theory states that the interpretation of utterances by humans relies crucially on the assumption that speakers intend their utterances to be relevant to the hearer. The original scope of the theory was linguistic pragmatics, but there have been attempts by its authors to extend it to general cognitive functions

(Sperber & Wilson, 1996; see also application of Relevance Theory to the Wason Selection task in Girotto, Kemmermeier, Sperber, & van der Henst, 2001).[26]

While relevance of stimuli to tasks has not been extensively addressed at a theoretical level in visual attention (with the exceptions discussed at the end of this section), many experimenters have found evidence of relevance-based mechanisms in attention, as discussed in Section 4.9. Experiments 6 and 7 (Chapter 4) also gave evidence of attention being directed to elements in the display most relevant to both the task and the way the participant carried out the task. It was further noted that the phenomenon of cognitive tunnelling occurs mainly when pilots do not expect certain events to occur in a particular part of the display, and thus do not seem to consider the region relevant to their task. Another example of relevance-based phenomena from aviation psychology is the fact that certain forms of scene-linked symbology reduce cognitive tunnelling only when the symbology is relevant to the task (Shelden, Foyle, & McCann, 1997).

The results from aviation psychology are particularly interesting, for the standard object-based attention theory has been invoked to explain them. There is nothing in the object-based attention theory as it is currently formulated that would make it sensitive to relevance-based effects. Gestalt grouping is not supposed to involve conceptual or conative information. Furthermore, the object-based explanation of tunnelling relies on

---

[26] There is not the space here to consider the explanatory and empirical adequacy of Relevance Theory. Suffice it to say that the enthusiasm that has greeted this theory illustrates the degree to which current thought in cognitive science is concerned with the relevance-based aspects of cognition.

an explicit formulation of the theory that is at odds with the data. Researchers in the aviation psychology community explain object-based attention in the following way: (1) visual stimuli are organized into Gestalt groupings, and (2) attention can only select one of those groupings at a time. Thus, if a stimulus appears while attention is allotted to another object, that stimulus will not be attended, and thus not perceived, unless attention is shifted to it. In cognitive tunnelling, that shift supposedly fails to occur. However, other experiments have given evidence for object-based attention even under conditions which would preclude the shifting of attention (Lavie & Driver, 1996; Experiments 5, 6, and 7, Chapter 4 above). Thus, cognitive tunnelling relies on an interpretation of the object-based thesis involving shifts of attention between objects, which is unfounded, and on some unexplained relevance-based mechanism that would lock attention onto the "relevant" stimulus in the first place.

More generally, the object-based thesis has no way of explaining the relevance-based phenomena discussed earlier. Perceptual organization, according to object-based theory, should not involve relevance-sensitive factors, and saying that the attentional selection of Gestalt groups can be influenced by the relevance of groups, or features thereof, to an observer's particular circumstances merely shifts the explanatory burden onto mechanisms of attentional control. Similarly, on Vecera's (2000) account of object-based attention augmented by spatial selection, it could be argued that the biasing of perceptual organization by spatial attention could happen in a relevance-based manner; however, this again merely shifts the explanatory burden onto the control of (spatial) attention. While many experimental data on attentional control have been generated (see review in Wright & Ward, 1998), there is still no adequate explanation as to how these mechanisms direct attention onto relevant or meaningful stimuli that does not invoke some form of homunculus.

There are few theories of visual processing which try to explicitly address the effects of relevance-based or action-related factors on attention. Two of the most notable ones are Bundesen's Theory of Visual Attention (TVA; 1990) and Hommel's Theory of Event Coding (TEC; Hommel, Müsseler, Aschersleben, & Prinz, 2001). Bundesen follows Broadbent (1971) in assuming that visual attention must perform two functions: *filtering* (which refers to selection of perceptual elements for further processing) and *pigeonholing* (which refers to the classification of items selected by filtering for appropriate responses). According to Bundesen's TVA, all elements in a display are processed in parallel. Selection for conscious perception (which, on the TVA, corresponds simultaneously to filtering and pigeonholing) occurs when visual elements are stored in short-term visual memory (STVM). TVA models this selection process as a horse-race: the elements that are processed most quickly make it into the STVM, until its capacity is reached, and the speed with which elements are processed depends on attentional parameters. A *pertinence* parameter captures the degree to which a visual element matches a perceptual category stored in long-term memory that is relevant to a *task*. Thus the pertinence parameter influences speed of filtering. A *perceptual bias* parameter determined the degree to which a visual element matches a perceptual category that is relevant to a particular *response*, thus reflecting speed of pigeonholing. Both of these parameters determine the rate of decay of the exponential parameters which model the processing speed of individual elements, thus showing how the probability of storage of each element in STVM is a factor of both filtering and pigeonholing.

On the assumption that the pertinence and perceptual bias parameters reflect what I have called relevance-based and stimulus-driven factors respectively, it could be claimed that the TVA is a mathematical model of how bottom-up and top-down factors influence attentional selection. However, it is not clear what pertinence and perceptual bias actually represent in Bundesen's theory. In Bundesen (1990), both pertinence and bias are defined in terms of perceptual categories stored in long-term memory, without defining them further. It is far from clear whether these 'perceptual categories' in fact reflect anything resembling conceptual or conative factors. Furthermore, the pertinence and bias parameters must be set by the visual system dynamically, as a function of task demands and observer's knowledge and interest, for attentional selection to occur, but Bundesen gives no account of how this might happen. In fact, he specifically claims that the goal of TVA is not to explicate the mechanisms by which top-down factors might affect attentional parameters, but rather to merely provide top-down mechanisms with parameters through which attention can be influenced: "No attempt is made to discard the notion that attentional selection is controlled by an intelligent agent, but a serious attempt is made to relieve the burden on the agent by placing a powerful mechanism at its disposal" (Bundesen, 1990, p. 523). Finally, Bundesen does not discuss the nature of the visual elements that the TVA places into STVM, but as was extensively discussed in Chapter 3, the issue of what gets selected by attention (perceptual groups vs. simple features vs. unprocessed regions of visual space) is crucial to providing a complete theory of attention. Bundesen’s TVA therefore amounts to a mathematical paraphrase of the statement “attention selects stimuli based on perceptual and relevance-based factors” without explaining what this means or how this might happen.[27]

The Theory of Event Coding (TEC; Hommel et al., 2001) also purports to be a general theory of how perception (with a special focus on vision) is influenced by both bottom-up and top-down factors. The basic premise of TEC is that the cognitive system encodes perceptual information and action plans in a common representational medium. Thus, perceiving a glass and the set of movements required to grasp the glass are, according to the TEC, both specified within a common representational format, called ‘feature codes.’ Feature codes are supposed to represent distal features of the environment (i.e., not the immediate stimulation at sensory transducers), which in theory are adequate for both the perceptual representation of elements in the environment and for the high-level specification of movements and actions. Sets of features are integrated into whole entities, called event codes, and which serve both as perceptual representations of events in the environment, and plans for executing actions, depending on whether a person is passively perceiving something or trying to act upon the environment. This integration is assumed to be performed by attention, and once a set of features is integrated into a particular event code, all other possible event codes in which the feature codes could participate are suppressed by attentional mechanisms.

---

[27] It should be noted that Bundesen’s TVA could probably serve as a mathematical model for attentional selection as performed by the MSCS mechanisms in the conative model, as long as it could be shown that the latter can generate the former, and how. This is beyond the scope of the present work, but represents an interesting line of future research.

Since event codes and the feature codes that they are composed of are influenced by, and support, specific actions, they provide a mechanism that can account for the effects of conative factors on visual perception. Thus, TEC is congenial to the conative model developed above. However, on the TEC view, as in Treisman's feature integration theories (Treisman 1988, 1998; Treisman & Gelade, 1980), attention is merely a "glue" that holds features together, in this case in a unified event code. Thus, TEC does not provide an account of how attentional selection itself is directed by conative factors. It could be argued that TEC provides a way of extending the standard object-based account of attention to include conative factors: on the assumption that the objects selected by attention are in fact event codes made up of feature codes, the objects that attention selects are shaped by conative factors. This would be consistent with the proposal that attention is object based because it selects features by binding them into object files (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992). However, as discussed extensively above, current accounts of perceptual organization propose basic visual features which have nothing to do with action plans (edges, surfaces, etc.) as the "ingredients" for perceptual organization. TEC provides no account of how perceptual organization, as it is currently understood, can be reconciled with feature codes. Indeed, given what is known about perceptual organization, it is not clear what the 'distal features of the environment' correspond to. If these correspond to high-level, abstracted features of the environment analogous to the affordances of ecological optics (Gibson, 1979), as the authors of TEC imply, then TEC does not provide an account of perceptual organization and selection of visual objects that is compatible with the state of the art on perceptual organization.

In addition to the problems raised by the feature and event codes of TEC, Hommel et al.’s (2001) theory, suffers from a problem similar to one encountered with Bundesen’s TVA (1990): TEC states that intentions and goals influence how attention integrates features into event codes, but provides no account of how this occurs. That is, both TEC and TVA state *that* attention must display relevance-based features, but do not propose how this might occur. They do not address the issue of how the visual system must be organized, at a functional level, so that relevance-based attention is possible; they merely state that the visual system must somehow be so structured. The conative model I propose here takes the extra step of exploring a possible functional architecture for vision that would allow for relevance-based selection.

On the conative model of attention, there is a natural explanation for how attention can be relevance-based. The MSCS mechanisms of perceptual organization feed partial patterns to visual attention. Attention then passes these partial patterns to the conative and conceptual MSCS systems. Any active patterns (partial or full) in those systems will interact with the partial patterns from perceptual organization, also through a MSCS mechanism. That is, the active patterns from the conative and conceptual systems will strengthen partial patterns from perceptual organization that are congruent with them (i.e., relevant to them), and inhibit incongruent perceptual patterns. It is this biasing of partial perceptual patterns by active top-down patterns that constitutes attentional selection and emphasis in the conative model. The fact that the top-down patterns are

themselves subject to MSCS mechanisms ensures that only a relatively small number of robust top-down patterns will be involved in attentional selection. The fact that partial top-down patterns can be used makes the selection process rapid. When no top-down patterns are dominant, they provide little biasing of partial perceptual patterns, thus resulting in bottom-up object-based attention, as commonly observed in the research literature. This is the default behaviour of the visual attentional system. When there are dominant top-down factors, especially cona
ive factors related to action and interests, the activation of partial perceptual patterns is biased in a relevance-based manner. This integration of features from early vision in the service of action, via the MSCS mechanisms of the conative model, provides a more natural account for the mutual dependence of action and perception that the event codes in Hommel et al.’s (2001) TEC.

5.2.3 The Case of Inattentional Blindness

One attentional phenomenon briefly mentioned above deserves a closer look. Inattentional Blindness (IB) is an interesting phenomenon for the conative model. This is because the explanation of the attentional mechanisms underlying IB given by its discoverers, Arien Mack and Irving Rock, is at odds with my model, but, as I will argue shortly, the experimental results which revealed the phenomenon of IB are compatible with the relevance-based aspects of the conative model. I will show that the conative model is a better explanation for IB than the account proposed by Mack and Rock (1998).

The claim that Mack and Rock make is that no conscious perception of visual stimuli occurs prior to attention being directed to the stimulus. However, according to the authors, this does not mean that stimuli are not processed without attention. In fact, they claim that stimuli are extensively processed, even to the point of being semantically interpreted, before they are selected for conscious perception by attention.

Mack and Rock's (1998) claim is based on the results of a methodology they first developed in order to investigate stimulus features that capture attention automatically. If a stimulus is to capture attention, it must be perceptible without attention. Following this assumption, Mack and Rock developed a dual-task paradigm. Participants were shown a small number of trials (from 6 to 8) wherein they were presented a fixation point for 1500 ms, a stimulus (a large cross) for 200 ms, and a noise mask for 500 ms. One arm of the cross was slightly longer, and participants were instructed to identify the longer cross. On one trial (usually the third or fourth one), in addition to the cross, another stimulus was shown. In some experiments, the stimulus was a small shape 2.3° away from the centre of the cross. In others, the stimulus was a texture or a Gestalt grouping pattern filling a circular area centred on the cross and whose diameter exceeded the length of the arms of the cross. Participants were not warned of the presence of this stimulus before it was shown. After that trial, participants were asked whether they saw the unexpected stimulus, and if so, what attributes of the stimulus (shape, colour, grouping, etc.) they noticed. In some cases, participants were asked to identify the stimulus that was shown from a menu of possible stimuli, even if they had not reported seeing the stimulus.

As the main purpose of the experiment was to determine whether participants detected the unexpected stimulus, the trial where this stimulus was first shown was called the ‘critical trial’, and the task involving the cross was termed the distractor task. After the critical trial, participants were again shown only the cross for a few (2 or 3 trials) and were instructed to compare the length of its arms again. After these additional distractor trials, participants were again shown the cross and the critical stimulus. In some experiments, participants were told to compare the arms of the cross and identify the critical stimulus. This was termed the ‘divided attention’ trial. In other variants of the experiment, participants were told to ignore the cross and only identify the critical stimulus. This was termed the ‘full attention’ trial. Some of Mack and Rock’s experiments involved both a divided attention trial and a full attention trial.

Many experiments following this general pattern were carried out, using a variety of critical stimulus features: shape, colour, location, perceptual grouping, apparent motion, words, and iconic faces (i.e., little happy-face figures). The overall result was that participants failed to detect the critical stimulus or any of its features at a rate of 75% on average during the critical trial. However, during the divided attention trials, detection and identification of the critical stimulus rose to about 75% (a reversal of the critical trial results), and on the full attention trial, identification was usually near 100%. Since participants’ attention was presumably engaged in the distractor task during the critical trial, and since the critical stimulus was perceived much more successfully during the divided and full attention trials than in the critical trial, it was assumed that participants were “blind” to the stimulus because they did not direct attention to it. Thus, the phenomenon was called ‘Inattentional Blindness.’

While Mack and Rock (1998) found that IB obtained for almost all stimuli, there were two types of stimuli that seemed immune to IB. One was the participant’s given name, the other was iconic facial expressions (e.g., ☺). These stimuli were identified about 90% of the time on critical trials, whereas other stimuli yielded identification rates of about 25%. Furthermore, stimuli that were visually similar to these, but which had different meanings (another word that differed from the participant’s given name by only one or two letters, or iconic faces displaying emotions other than happiness) yielded identification rates that were closer to (but nevertheless higher than) the 25% produced with other stimuli. On the basis of these results, Mack and Rock proposed that all stimuli are “deeply” processed, often to the point of semantic interpretation, independently of attentional selection, but they are consciously perceived only when they are attended. Furthermore, only semantically “meaningful” stimuli seem to be able to capture attention that is already otherwise engaged.

The fact that participants attend more readily to their own name rather than geometric shapes or grouping patterns is fully consonant with the relevance-based aspects of the conative model, as discussed above. However, the interpretation of these data on a conative model differs from Mack and Rock’s. On the conative model, processing visual stimuli beyond the level of edges and lines requires some degree of attention. The activation of partial patterns and MSCS mechanisms allow attention to be quickly deployed in a flexible and context-dependent manner so as to favour the processing of relevant information. On this model, attentional selection is neither “early” nor “late”, but can operate at various processing levels flexibly, as required by the context of an observer. Mack and Rock (1998), on the other hand, suggest that all visual stimuli are processed without attention, and that only conscious awareness of stimuli requires attention. Thus, Mack and Rock espouse a late-selection view of attention.

The position that attentional selection is purely “late,” i.e. that is occurs solely after semantic interpretation of stimuli, is problematic (Pashler, 1998). While some evidence supports late-selection models, as Mack and Rock’s (1998) data demonstrate, there are also many results that indicate that attention is a prerequisite for semantic interpretation of stimuli (e.g., Treisman’s feature integration model). Thus a purely late-selection model cannot account for a significant body of experimental data.

More problematic, perhaps, is the criterion Mack and Rock use for determining whether a stimulus is attended. They relied on the assumption that only conscious perception reflects attentional selection. This assumption, in turn, relies on the fact that only stimuli with high semantic relevance seem immune to IB. The interpretation of this fact depends crucially on whether it can be said with certainty that attention was completely allocated to the cross, and to no other part of the display at all, during the critical trials. Mack and Rock certainly think so, as they explicitly state that stimuli must capture attention away from the cross to be perceived. Thus, they clearly assume that attention can be completely withdrawn from everything except the cross in the critical trial. Yet they have no evidence to that effect.

Mack and Rock (1998) note that in their early studies, their results suggested that colour, location and numerosity were immune to IB. They later realized that when critical stimuli fell within a disc defined by the longest arm of the cross, as they did in their early experiments, hardly any IB occurred, but when stimuli fell beyond the cross, as in their experiments using Gestalt patterns that covered an area larger than the cross, or in later experiments where the length of the longest arm of the cross was reduced from 4º to 0.6º of visual angle, considerable IB occurred. This is consistent with the zoom-lens variant of the spotlight model, where the attentional spotlight is defined by the arms of the cross. In any case, these results clearly show that on many critical trials attention was divided between the cross and the critical stimulus. Thus, it is difficult to claim with certainty that absolutely no attention was allocated to the critical stimulus during critical trials.

The claim that all critical stimuli were processed, but only stimuli that were consciously perceived were attended to, is contradicted by research on the role of attention and consciousness in action. It is known that attention to stimuli relevant to a task plays a large role in the motor control required to carry out specific actions (Cisek & Turgeon, 1999; Craighero, Fadiga, Umiltà, & Rizzolati, 1999; Milner & Goodale, 1998). Yet Milner and Goodale have also shown that stimulus features that are required to guide action do not have to be consciously perceived. In fact, many of their brain lesion patients are able to execute movements for which they must devote visual attention to a particular stimulus, but nevertheless have no conscious access to the features of the stimulus that guide their movements (Goodale & Milner, 1992; Milner & Goodale, 1998). Furthermore, many of their patients have lost their ability to execute movements guided by features of visual stimuli, and yet have full conscious awareness of the features that would normally guide their movements. Clearly, the conscious perception of a stimulus (or lack thereof), and the ability to report conscious perception of the same stimulus (or lack thereof), cannot be used as a criterion to establish whether the processing of the stimulus involved attention, as Hardcastle points out (2003).

While the phenomenon of IB is real enough, Mack and Rock's explanation of it has serious problems. It equates consciousness and attention, it assumes a late selection model of attention, and it makes it impossible to provide a parsimonious account of the role of attention in visually-guided action. The conative model, on the other hand, can explain IB while avoiding these difficulties. On the conative model, attention is directed to the perceptual structures that are most relevant to a person in a given situation. Note, however, that while the relevance of a stimulus can be determined by the task a person is engaged in, certain types of information can be universally relevant. For instance, a person's given name is highly relevant to them, as the now-classic cocktail party effect shows (Moray, 1959). Facial expressions also have high ethological value. On the conative model, then, attention is re-deployed during each trial of the IB experiments. During the critical trial, only the cross is relevant to the task, and has been made so by the two or three regular trials preceding it. Therefore, most attention is allocated to the cross, but a smaller amount is also allocated to the critical stimulus, resulting in the high levels of IB found for those trials. When the critical stimulus is the participant's name or an iconic facial expression, however, both the cross and the critical stimulus are relevant, the cross because of the task, and the name or happy face because of their more generalized relevance. Attention is therefore more evenly divided between the cross and the critical stimulus, resulting in the much lower levels of IB found by Mack and Rock. During the divided attention trials, the context of the experiment itself makes the cross and the critical stimulus relevant, and thus are both attended more symmetrically than in the critical trial.

On the conative model, we do not have to follow Mack and Rock's suggestion that the critical stimulus was processed but not attended during the critical trial, except when the critical stimulus was the participant's name or a facial expression. Rather, we can assume that attention was asymmetrically divided between two stimuli, as a function of their relevance to the participant given that participant's context. This allows us to explain IB without assuming that attention selects stimuli only after they are semantically interpreted, and that only consciously perceived stimuli have received attention. It also allows us to account for IB while retaining the ability to explain a role for attention in visually guiding action.

5.2.4 Spatial Attention in the Conative Model

As mentioned above, there have been a number of attempts to integrate spatial and object-based visual attention. I argued that these attempts do not adequately account for data that indicate (1) that attention and perceptual organization governed by inference to the best explanation (IBE) are highly interpenetrated and (2) that attention operates in a relevance-based manner, but that the conative model does account for these aspects of visual attention. I will now argue that, by accounting for these two aspects, the conative model also explains how attention can seem object-based under certain conditions, and space-based under others.

First, let us note that the original intent of the claim that attention selects Gestalt groups rather than regions of the visual field was to argue that, at least under certain conditions, attention selects stimuli independently of their spatial distribution. This claim is internally inconsistent: one of the Gestalt grouping principles, grouping by proximity, is inherently spatial. It would *prima facie* provide a rationale for uniting spatial and object-based attention. Oddly enough, there is only one model of attention which exploits this fact, Logan's COntour DEtector (CODE) theory of visual attention (1996). The CODE theory can account for the fact that the processing cost of attending to two distinct targets increases as the spatial separation between them increases (Eriksen & Eriksen, 1974; Posner, 1980), on the assumption that increased spatial separation is equivalent to weaker grouping. However, grouping by proximity does not exhaust perceptual grouping. Thus, it is difficult on the CODE model to account for the fact that targets can be grouped by motion (Driver & Baylis, 1989), or that targets that are close together but displayed on separate objects are responded to more slowly than more distant targets located on the same object (Lavie & Driver, 1996).

As the principle of grouping by proximity implies, objects inherently have spatial extent. Thus, object-based accounts should be able to explain spatial effects as well as non-spatial object-based effects. For this, object-based theories of attention must rely on accounts of perceptual organization that explain how and under which conditions spatial factors contribute to perceptual organization. As the Gestalt principles provide no basis for saying when one principle rather than another is active in a complex scene, they are not adequate for our purposes; but IBE-based perceptual organization provides a good basis for perceptual organization that is sensitive to spatial distributions. If spatial distributions provide the best explanation for spatial structure, i.e., if they provide the most generic explanation of the stimulus, consistent with the rules of physics as encoded into our visual systems by evolution, then spatial factors will dominate perceptual organization and shape a-objects. If increasing spatial separation makes this organization difficult, this will be reflected in more difficult attentional selection, and less efficient responses, as attention and perceptual organization are intertwined.

The relevance-based aspects of attention also provide a rationale for integrating object-based and spatial factors in visual attention. Attention will select and emphasize stimuli based on the properties of a stimulus that are relevant in a given situation. Thus, if spatial factors are most relevant to a task, they will dominate attentional distribution; but if spatially invariant object-based factors are most relevant, they will dominate. Furthermore, on this account, if both spatial and object-based factors are important to a task, then they will likely both play a role in allocating attention.

Let us see how relevance-based mechanisms fare in explaining spatial factors in existing research. One of the first studies to establish a spatial component to attention was Eriksen and Eriksen’s 1974 study involving the flanker task. In that task, a participant is instructed to classify a letter in a display based on certain features – typically, participants respond one way if the letter contains curved lines, and another way if the letter contains only straight lines. The display usually contains a row of five letters, the middle one being the target stimulus, and the flanking letters being distractors (hence the 'flanker task'). The distractor letters belong either to the same category as the target letter (congruent distractors) or to the other one (incongruent distractors). Eriksen and Eriksen found that congruent distractors facilitate responses to the target (reducing response times) and that incongruent distractors inhibit responses to the target. Furthermore, they found that the effect was strongest for distractors directly adjacent to the target; the more distant congruent distractors enhanced responses less than nearer ones, and the same for incongruent distractors, *mutatis mutandis*.

The Eriksen and Erisken (1974) task in principle involves only the target stimulus; in theory, the flankers are irrelevant. However, participants were not instructed to ignore the flankers, and neither did task demands compel participants to ignore the flankers. Thus, on the conative account, there was no dominant top-down factor to bias attention. Under such conditions, bottom-up factors of perceptual organization should dominate. And, under these conditions, only grouping by similarity (congruence/incongruence of the flankers) and flanker distance provide a basis for organizing visual stimuli. These two factors are clearly reflected in the response patterns observed by the experimenters.

On the conative model, if task demands make spatial factors more relevant than object-based ones, spatial factors should dominate, and even override, constraints on attentional allocation from spatially-independent perceptual organization. There are data to support this claim. Using the two-dashed-line "X"-shaped stimulus discussed in Chapter 3, Lavie and Driver (1996) found that participants respond to targets shown on the same line but separated by 8.1º of visual angle faster than to targets separated by only 3.1º but located on separate lines, thus establishing a single-object advantage for their stimulus. In one experiment, Lavie and Driver seemingly eliminated this effect by using visual cues to emphasize one side of the display (left or right) before the targets were shown. Targets appeared near the cues (and were thus displayed near to each other but on separate lines) on 70% of the trials ("near-valid" condition). During the remaining trials, targets appeared in three possible configurations: near to each other but on different lines and on the side opposite the cued side ("near-invalid" condition; 10% of the trials), on different lines but separated by 7.8º ("far" condition; 10% of trials), or on the same line ("object" condition; 10% of the trials). Under these conditions, targets appearing in the "near-invalid" condition were responded to fastest.

Lavie and Driver (1996) noted that trials for the conditions where the cues were invalid ("near-invalid," "object" and "far"), the order of response times followed the order of spatial separation between targets. Thus, for uncued target locations, participants responded most rapidly to targets in the "near-invalid" condition (3.1º apart), somewhat less rapidly for "far" targets (7.8º apart), and least rapidly for targets in the "object" condition (8.1º apart). The difference in reaction times between targets in the "near-invalid" and the other two uncued conditions was significant; however, the difference between "far" and "object" targets was not. Based on these findings, and on the fact that cued targets were responded to significantly faster than uncued ones (even "near-invalid" ones), Lavie and Driver proposed that object-based attention operates only once a spatial attentional "spotlight" is established. They further suggested that the cues used in the experiment disrupted object-based mechanisms, thus making only spatial selection effective.

Lavie and Driver's (1996) proposal that object-based attention operates only once spatial attention has selected a region of space has the disadvantage of requiring an account of the conditions under which object-based mechanisms might be disrupted or suspended within the spatial spotlight (of which Lavie and Driver provide none). The conative model, however, provides a natural account of why visual attention might cease to behave in an object-based manner and become more sensitive to spatial features of a display under certain circumstances. The cues, which predicted target location 70% of the time, establish a region of "maximum relevance" around them, and thus participants direct most attention to the target locations nearest the cues. When targets appear outside of the cued region, the visual system has to initiate a scan of the display in order to find and compare both targets. Under these conditions, organizing the display according to spatial proximity rather than objects is more useful to the target comparison task, and thus target separations become more relevant than objects for comparing the uncued targets.

Based on the analyses of the Eriksen and Eriksen (1974) data and Lavie and Driver (1996) results above, we see that the relevance-based mechanisms and IBE-based perceptual organization of the conative model provide natural accounts of the allegedly spatial attentional selection observed by Eriksen & Eriksen, and Lavie and Driver, while *also* accounting for object-based attention, and the relevance-based phenomena discussed in Sections 4.9 and 5.2 above, including the phenomenon of Inattentional Blindness. The conative model therefore offers the opportunity of integrating many more aspects of visual attention in an empirically testable framework than have heretofore been accounted for by current theories of visual attention.

5.3 Can The Distribution of Attention in a Scene Be Predicted?

An important goal of any theory of visual attention should be to predict where observers direct attention in a visual scene, in addition to explaining where and why attention was allocated in previous experiments. On existing theories of visual attention, predicting the allocation of attention seems simple enough. On the spotlight theory, attention is allocated mainly to a roughly circular region of space surrounding a stimulus that has attracted attention. On the object-based theory, attention will be allocated primarily to one object (i.e., one a-object, corresponding to a concrete or virtual object) that is the focus of attention. We have seen, however, that these models fail to account for a number of experimental observations that I called ‘relevance-based’ above, and leave the burden of explaining why some stimulus (cue or object) attracted attention in the first place. Indeed, it is not clear how the object-based or spotlight models (alone, or even in combination) could be used to predict what a person is attending to in must real-life situations.

Does the conative model fare any better? The model was used to give a unified account of object-based, space-based and relevance-based aspects of attention, including Inattentional Blindness, in the previous section. But retrospective fits to data do not ensure that the model has predictive power. Furthermore, if the conative model is more-or-less correct, then the distribution of attention depends very much on the vagaries of the task in the context of which attention is being deployed, as well as on the state of the individual performing a task. Thus, predicting where attention is deployed using the conative account seems difficult *prima facie*.

I will show below how the conative model can be used to predict how attention is deployed in a given situation. First, however, two caveats. The first is that if the conative model is correct, the distribution of attention in a given scene can be predicted only if both the cues for perceptual organization in the stimulus and the context of the task for which a person is attending to a scene are first analyzed. Secondly, as task context depends very much on aspects of a person that might not be accessible by other people (e.g., background knowledge, expectations, goals), it might be the case that, as a general principle, the deployment of attention can be predicted only approximately.

As discussed above, attention is directed by three systems of constraints in the conative model: stimulus-driven perceptual organization (including the rules of perceptual organization as discussed in Section 5.1), top-down conceptual factors, and conative factors (goals, expectations, needs). The dynamic satisfaction of these three systems of constraints directs attention. Clearly, then, to predict where attention will be deployed, one must determine which systems are likely to dominate the visual system in a given situation. Three general classes of situations can be distinguished: cases where there are no dominant constraints in the top-down (conceptual or conative) systems, cases where a small number of constraints dominate the top-down systems and these constraints are congruent with the bottom-up constraints, and finally cases where the top-down and the bottom-up systems are in conflict.

In cases where no constraints in the conative and conceptual systems are dominant, the net biasing effect of the top-down constraints should be roughly nil. In those cases, the constraints from stimulus-driven perceptual grouping can dominate how attention gets deployed. Attention would behave in a manner resembling the object-based behaviour observed in many experiments. Or, if the only stimuli are cues as in the spatial cueing paradigms used by Posner (1980), attention would appear to be spatially based. In either case, attention would be controlled by properties of the stimulus. There is research (Feldman, 1999; Palmer & Rock, 1994) that suggests that complex scenes are organized hierarchically; that is, objects are themselves organized into larger sets or groupings, and objects themselves are perceptually articulated. On the conative model, where attention is not withdrawn onto a single object unless focused thus intentionally or by task demands, attention would emphasize (and thus facilitate the processing of) information available at the higher, more general levels of the perceptual hierarchy; elements at the lower levels of the hierarchy (individual objects in a complex naturalistic scene, object parts) would receive relatively less attention. These cases would correspond to situations where a person is observing a scene with no particular intent and is not required to perform a particular task using the scene, and no element of the scene is particularly meaningful to the person. Admittedly, this combination of factors is highly unlikely to obtain in the real world, but it can certainly be contrived in an experimental setting.

In cases where there are dominant constraints in the top-down systems, we should observe the relevance-driven deployment of attention discussed in Section 5.2 and observed in Experiments 2, 3, 4, 6 and 7 (Chapter 4). Thus, objects in a scene, or visual features of an object, that are relevant to a task will guide the allocation of attention. Mapelli, Cherubini and Umiltà (2002) observed that attention was distributed in an object-based fashion in their experiments only when global object properties were relevant to the task employed in the experiment. As discussed in Section 5.2.4, if spatial factors are most relevant to an experimental task, they will dominate attention. Note also that certain conative constraints are likely to be “harder” than others – i.e., they are more likely to be activated or dominate the system of conative constraints because of their ethological value, determined through evolution or experience. Thus, as noted in Section 5.2.3 on Inattentional Blindness, it is known that a person’s given name will attract their attention almost universally, both in visual (Mack & Rock, 1998) and in auditory form (Moray, 1959). I would predict that stimuli that are highly personally meaningful to a person (names or faces of loved ones, words associated with past traumas) would also attract attention in an almost universal manner. As suggested earlier, it might also be the case that a sensitivity to certain environmental stimuli (e.g., objects hurtling towards one’s head, fire or large flashes of light, human faces) has been “hard-wired” into the conative system, thereby guaranteeing their almost universal ability to attract attention. As noted above, Easterbrook (1959) observed that under stressful situations, people focus increasingly on stimuli which carry threat information, thus also suggesting an effect of the congruence between conative and stimulus-driven constraints on attention. Thus, it would be safe enough to predict that visual attention will naturally favour stimuli congruent with the task or existential context at hand.

On the conative model, top-down influences, such as task demands and goals, bias attentional selection towards perceptual structures or features that are consistent with the top-down constraints. What happens when the top-down constraints are in conflict with the bottom-up ones? Could attention be directed towards visual stimuli so as to favour the processing of information relevant to a task despite the fact that the information appears in separate perceptual structures? Perhaps. On the conative account, if the three constraint systems are in conflict, the master MCSC process that integrates them would take longer to settle into a configuration that satisfies the constraints;[28] furthermore, the global solution to the constraint satisfaction problem posed by the conflicting constraints might be sub-optimal; that is, the pattern of attentional allocation might reflect a compromise between the constraints rather than an optimal solution where attention selects relevant stimuli effortlessly and automatically.

In practical terms, allocating attention while performing a task where perceptual and conative or conceptual cues conflict would be marked by an increased conscious effort to focus on relevant stimuli, and slower and less accurate performance. An indication that this might in fact be the case comes from the well-known Stroop task (Stroop, 1935). In the simplest version of the task, participants are shown names of colours printed in coloured ink; sometimes the ink and the name match, sometimes they

---

[28] In terms of perceptual organization based on inference to the best explanation, as discussed above, this corresponds to the visual system taking longer to make a “best inference” as to perceptual structure.

don’t. They are instructed to either name the colour of the ink or read the printed colour name. In every case, there are two perceptual cues (colour word and ink colour), which can be either consistent or in conflict with the naming task. When the ink and the word don’t match, responses are always slower and less accurate than when they do, thus already suggesting that attention is engaged in resolving a constraint conflict. Also, when the ink and the word don’t match, participants find it harder to name the ink colour than the colour word. While these results could be interpreted as the result of a purely bottom-up conflict (ink colour vs. word colour) coupled with the automaticity of reading processes (i.e., their overlearned nature), they also have a natural account on the conative model. When the ink and the colour word don’t match, there is always a conflict between the constraints from the task (naming the appropriate colour) and the constraints from the stimulus. Thus, performance is automatically more laborious and less efficient than in the case where ink and word match. The overlearned nature of reading makes all words more meaningful and “relevant” than colours *to the task of naming words in response to visual stimuli*. This ensures that naming the ink will be more difficult than naming the word – the constraint produced by the intention to read is stronger than the constraint produced by the intention to name an ink colour, and thus it is easier for the visual system to (eventually) resolve the conflict between stimulus and task goal in favour of the goal when the goal is to read a word rather than name a visual feature.

Reading is a highly overlearned task, and thus, is seems it is easier for attention to override the cue from ink colour and focus on the word rather than the other way around. In the cases where the task at hand is not so overlearned, attempting to direct attention to stimuli that are not perceptually grouped or that do not correspond to a particular visual feature is likely to be that much more effortful and problematic; should that effort falter, attention is likely to revert to more automatic modes of functioning such as focusing on object or relevant perceptual groups. Again, this is seen in the cases where participants in the Stroop task name the colour word when they are supposed to name the ink. On the conative model, we can predict that as participants gain experience with a task that requires them to integrate information across perceptual structures, the deployment of attention will move from effortful and inefficient to efficient and automatic, thus eventually overriding constraints from perceptual organization. In other words, I predict that with training, people can learn to attend to stimuli as though they were within a single object even when they are not. There is some anecdotal support for this notion: pilots report that in tasks where they are required to cross-check many instruments to perform a particular task, they initially use an effortful visual scanning pattern that inspects each relevant instrument sequentially, as though each object gave rise to individual a-objects. As they gain experience with the tasks, pilots report they eventually learn to attend to the various gauges as a whole and extract the relevant information from them without inspecting them serially. It is as though the set of gauges were one object *for the purposes of the task*, giving rise to a single a-object (likely with articulated sub-parts).

I predict that the phenomenon of learning to attend to information from many objects as though they formed a whole, thus overriding the perceptual structure given by stimulus properties, is a general one. That is, people can probably learn to attend to disparate stimuli as though they were displayed in a single object as long as the conjunction of stimuli is relevant to a task. This ability would come at the cost of much training, during which the allocation of attention would require conscious effort – a phenomenon that we call "concentrating" when we notice it in our everyday lives.

6. ATTENTION IN THE WILD: THE CASE OF HUDS

The conative model developed in the previous chapter was conceived to capture how attention is actively deployed in the course of purposeful human engagement within the context of a particular environment and task – the "life worlds" (Heelan, 1983) or *Umwelten*[29] of the phenomenological tradition in philosophy. The model must therefore be examined in the light of the explanatory power and methodological tools it brings to the study of real-life situations. I have chosen to study HUD use in cockpits as a case study of attentional behaviour "in the wild," to borrow an expression from Hutchins (1995). I present general methodological implications of the conative model for the study of HUD use and design in general. I also examine what the conative model contributes to the specific issues of cognitive tunnelling and mixed frames of reference in HUDs, two issues which motivated much of the experimental work I carried out (Chapter 4).

6.1 HUD Research in General

Human-factors or ergonomics-based studies of HUDs tend to take a "usability" approach. To put it crudely, usability studies of computer interfaces often focus largely on the degree to which particular design features of an interface affect a user's subjective sense of comfort and ease of use (Ravden & Johnson, 1989; though see Norman &

---

[29] While the term is now associated with the phenomenological tradition, A. Clark (1997) attributes the term to the biologist/philosopher J. von Uexküll, who noted that, functionally speaking, every organism lives in its own world, or *Umwelt.*

Draper, 1986, for a more cognitively-oriented take on usability). In terms of research on HUDs, one often sees studies where design features (colour, number of symbols, etc.) are manipulated as independent variables, subjective “usability” ratings are used as dependent measures, and little to no thought is given to the role that cognitive or other user-dependent factors might play in mediating the relationship between design features and usability ratings, or to whether usability ratings actually reveal anything about how well a user will perform a task using the HUD (Sarter & Woods, 1991). A relative step forward has been taken recently by researchers who have attempted to make use of psychological constructs to explain the performance of pilots flying HUD-equipped aircraft. One example is the construct of “Situation Awareness” (SA; Endsley, 1995), proposed as a mediating factor between HUD design and performance, although SA metrics tend to rely almost exclusively on subjective reports, and the construct of SA itself has serious conceptual problems (see Sarter & Woods, 1991). It is well-known that behaviour is often determined by factors that people cannot verbally report (Milner & Goodale, 1998; see also a classic review in the area of social psychology by Nisbett & Wilson, 1977), and a recent study of other cockpit technologies has shown that subjective performance ratings can in fact be at odds with objective measures of performance quality (Herdman et al., 2001b). Another example, discussed extensively above, is the use of object-based attention to study cognitive tunnelling.

Based on the conative model of attention I propose, and on my experience studying attention in an action-based setting, I propose that study of the human factors involved in HUD use needs to proceed on the assumption that HUDs are *tools* used for the purpose of carrying out *specific activities*. All research on HUDs needs to consider four issues: the HUD design, the task(s) for which the HUD is used, the person performing the task, and the context in which the task is performed. HUD design features need to be examined both in light of what is known about human cognition in general, a part of which is knowledge of the human attentional system, and as they relate to the task context.

A particular HUD design decision will have different implications depending on the task performed with the HUD and the conditions under which it is performed (high stress, low stress, etc.). Some of the research previously cited indeed suggests that the benefits of scene-linked symbology might depend on the relevance of the scene-linked symbology to the task (Shelden et al., 1997). Also, individual differences such as extent of training or fatigue can have dramatic effects on observed performance, which might not be detected with subjective measures only. Experiments 1 through 4 in Chapter 4 demonstrate that different attentional strategies – which can be either learnt or taught – have a significant impact on performance, which suggests that some HUD usage issues might be the result of, or alleviated by, appropriate attentional strategies and training. Indeed, evidence for this is provided by the inside-out vs. outside-in debate in attitude indicator design. An aircraft's attitude is its general spatial orientation (pitch, roll) relative to a terrestrial frame of reference. Aircraft designed in North America tend to use attitude indicators where the aircraft symbol is stationary whereas the horizon symbol moves relative to the aircraft, mimicking the view a pilot would normally have of the real horizon from inside the cockpit (hence "inside-out"). Aircraft designed in the ex-Soviet

Union usually use attitude indicators where the artificial horizon is stationary while the aircraft symbol moves, representing a “bird’s-eye-view” of the aircraft flying over the earth (hence “outside-in”). Studies have shown that novice pilots (and Soviet-trained pilots) tend to find the outside-in design more intuitive and easier to use, whereas pilots with extensive experience using the inside-out design prefer it to the outside-in design (Cohen, Otakeno, Previc & Ercoline, 2001). Along similar lines, Experiment 7 suggests that the way a user organizes attention in a HUD might vary as a user gains experience with particular tasks; this would be consistent with the classic laboratory studies showing that expert chess players recognize and attend to global patterns differently than do novice players (Chase & Simon, 1973).

The fundamental methodological implications of the conative model of attention for HUD research are that: (1) the dependent measures in HUD studies should involve performance on a task where the user is required to manipulate something, either on the HUD itself, or the aircraft via the HUD; and (2) conative and other user-driven factors (e.g., training) should be independent variables manipulated in the study, and not treated as potential confounds to be “controlled for” and “factored out” in the design. The first desideratum – that HUD studies involve examining how HUD use mediates a task – is met easily enough, as most (if not all) HUD studies require participants to fly an aircraft using the HUD; the danger is in limiting outcome measures only to subjective reports and epistemic judgement tasks such as feature comparison tasks, though these measures do have their role to play. The second desideratum, that of integrating user-driven factors into the design of a study, is more problematic. Many HUD studies have relatively complex designs and adding further dimensions to the design can be difficult to accommodate, especially when those added dimensions are factors that are difficult to objectively quantify in the first place, such as motivation or experience-dependent attentional strategies. Thus, HUD research would benefit from integrating conative and conceptual factors into experimental research, from using a combination of subjective reports and quantitative performance measures.

6.2 Specific Issues: Cognitive Tunnelling & Mixed-FOR HUDs

Specific human factors issues with HUDs, namely cognitive tunnelling and the development of mixed-FOR HUDs, motivated much of the empirical and conceptual research that led to the development of the conative attention model. It is therefore worthwhile to examine what the conative model contributes to the study of tunnelling and mixed-FOR HUDs.

As noted earlier, cognitive tunnelling has been attributed to object-based attentional mechanisms. It has been assumed that (a) the HUD and the environment are each perceived as a perceptual object in its own right and (b) the visual system can therefore attend to either the HUD or the outside scene only at any given time. Thus, if a pilot's attention happens to be directed to the HUD, it is likely that information not on the HUD will be processed less efficiently (McCann, Foyle, & Johnston, 1993), and unexpected events in the outside scene might be missed altogether (Wickens & Long, 1995). Therefore, designing HUD with conformal or scene-linked symbology would reduce or eliminate tunnelling, as the HUD and the outside scene would therefore become perceptually fused (McCann & Foyle, 1995; Wickens & Long, 1995). In sum, tunnelling is seen as a design problem, in that the visual system is assumed to inevitably tunnel onto HUDs as they are usually designed, and it is seen as having a design solution, since the design of HUDs can be modified but the visual system's response to HUDs can't.

It is instructive to examine the assumptions underlying the object-based explanation of cognitive tunnelling in light of the conative model. First – do the HUD and the environment each constitute individual objects? On the conative model, the answer would be: if the global properties of the HUD or the outside scene (such as they may be), or relationships between the elements of the HUD (e.g., a relationship between torque, airspeed and altitude for helicopter HUDs) or the outside scene are most relevant aspects of the display to a given task, then they will be attended to as individual objects. However, that also means that it cannot be taken for granted that the HUD or the outside scene are in fact attended to as objects without first considering the context within which they are being used. It might be the case that, in the course of a particular task, some elements of the HUD that are particularly relevant to the task are perceived as individual objects, whereas the rest of the HUD is attended to as a whole (or as an inchoate mass of elements). This possibility could be the basis for a variation of the object-based explanation of cognitive tunnelling: the pilot becomes focused not on the HUD as a whole but on an individual object that is part of the HUD. Again, though, which particular element or perceptual groups within the HUD are being attended as objects at any given time depends on task demands and the pilot's experience. In brief, what parts of the HUD and the outside scene have been organized and are attended as objects – what the pilot might have tunnelled onto – is contextually determined.

Another contextually-determined phenomenon is the degree to which attention has been symmetrically deployed among the relevant objects. Wickens and Long (1995; see also Fadden, Ververs & Wickens, 2001) have recognized this in part with the observation that tunnelling in its extreme forms occurs only when unexpected events occur in areas away from the pilot's focus of attention. The conative model extends this observation by positing a continuum for divided attention. On one end, attention is divided between objects completely symmetrically. On the other, attention is entirely focused on a single object, as is said to be the case with cognitive tunnelling in HUDs and in the Inattentional Blindness studies of Mack and Rock (1998) reviewed above. On this view, the degree of asymmetry in attentional distribution is influenced by context and organism-driven factors in addition to being responsive to stimulus-driven factors (see discussion in Chapter 5).

The context-sensitive and relevance-based aspects of visual attention raise the possibility that pilots using HUDs could learn to avoid "tunnelling" onto a single object, e.g., by deliberately cross-checking between the HUD and the outside scene at fixed intervals, even when there is no obvious reason to do so. Studies by Foyle, Dowell, & Hooey (2001; see also Dowell, Foyle, Hooey, & Williams, 2002) supports this possible solution to tunnelling: in these studies, the HUD (a numeric altitude indicator) was located either in the centre of the display, or in an off-centre position, and tunnelling occurred only with the centred HUD. In the first condition, the HUD and the centre of optic flow from the outside scene are superimposed, and since the task required constant monitoring of the HUD, it is possible that pilots took it for granted that they had easy access to the motion cues they needed to fly the aircraft as they focused on the HUD. The pilots would therefore be less motivated to actively attend to the outside scene, thereby inducing a HUD-outside scene attentional asymmetry in favour of the HUD. However, in the second condition, pilots had to shift their gaze and their attention between the HUD and the centre of optic flow in order to perform the task; they therefore could not take it for granted that they were aware of the necessary motion cues and had more incentive to deliberately attend both to the HUD and the outside scene more-or-less symmetrically. Such evidence, in combination with the experimental results reviewed in Chapter 4, suggest that attentional strategies, either explicitly learnt or induced by task demands, can be the cause of, as well as the solution to, cognitive tunnelling. On this view, conformal and scene-linked symbology might simply represent the use of perceptual cues to promote a strategy of distributing attention more symmetrically between two objects (the HUD and the outside scene). Thus, conformal or scene-linked symbols, by directly overlaying an element of the outside environment such as a runway, might attract attention both to the symbology and the environment. Further research is in order to explore whether conformal symbology is mainly a perceptual phenomenon or whether it also exploits the conative aspects of attention, as the answer to this question will affect the conditions under which conformal symbology is effective.

The study of mixed-FOR HUDs can also benefit from adopting a conative view of attention. Such HUDs have only recently been developed, and little is known yet about their usefulness. In one of the few existing studies of a mixed-FOR HUD, Herdman et al. (2001a), showed that pilot performance benefited from the use of a mixed-FOR HUD where three elements – airspeed, torque and altitude – were yoked to the pilot's head movement whereas the rest of the HUD symbology remained stationary relative to the aircraft. However, the mixed-FOR HUD was evaluated only within the context of a limited number of training tasks requiring pilots to extensively look off to the side (at a right angle from the aircraft's horizontal axis). The study did not examine whether pilot's attention to the aircraft-referenced parts of the HUD or to the outside scene was significantly compromised by the head-yoked elements. The results of Experiment 7 (Chapter 4), however, allow us to make some predictions about mixed-FOR HUDs: if pilots need to make use of the precise location or trajectory of the head-yoked elements for a given task, these elements will likely excessively attract pilots' attention until pilots become used to using them. In other words, in the course of certain tasks, novice pilots might well have difficulties monitoring the head-yoked symbology without "tunnelling" onto it. It is unclear to what degree simply interrogating these elements requires their precise trajectory to be recovered by the visual system as the task in Experiment 7 did, and this issue needs to be studied in controlled experiments. However, the "good news" is that the attentional difficulties induced by the visual system's compensation for self-motion should fade with experience. That being said, it should be noted that all the general issues concerning cognitive tunnelling with aircraft-stationary HUDs also apply to the head-yoked symbology of mixed-FOR HUDs. If attentional strategies, whether determined by training or by implicit task demands, make the head-yoked elements the most relevant elements of the display, pilots risk tunnelling their attention onto them at the expense of other aspects of the display. However, this effect could be counteracted by ensuring pilots have available to them appropriate attentional strategies and are given the appropriate incentives for distributing attention more symmetrically while still attending to relevant elements in the display.

Note that, on a standard object-based account, tunnelling onto head-yoked symbology would be mostly the result of stimulus-driven factors, and could only be mitigated through appropriate HUD design. The conative model not only allows us to predict potential problems with a new HUD technology, it also offers a wider range of solutions (design- and user-based) to HUD problems than does the standard object-based account.

7. GENERAL IMPLICATIONS FOR COGNITIVE SCIENCE

Recent work and current debates in the area of visual attention grow out of the standard information-processing approach to cognition. The information-processing (IP) paradigm had undergone many changes over the years and many of its aspects have been challenged (see discussions in Clark, 1997, and Thagard, 1996). Nevertheless, there are two theoretical commitments from the historical core of the information-processing paradigm that still inform much research in cognitive science. One is a focus on cognition as an exercise in constructing representations of the world from perceptual data. The other is what I call ‘serially modular’ cognitive architecture: the assumption that the cognitive system is divided into peripheral modules that process perceptual data in serial stages (Sternberg, 1969), and a central module where perceptual information is integrated and used for general cognition (Fodor, 1983).[30] In many cases, the experimental methodology used to study the perceptual aspects of cognition have been developed so as to investigate the serial construction of perceptual representations assumed to occur within the peripheral modules. Thus, reaction times, error rates and epistemic tasks in response to highly artificial stimuli have become the mainstay of much research on perceptual cognition. Conversely, work on more “central” types of cognition (problem solving, decision making, utterance interpretation) often focuses more (though not exclusively) on sampling participants’ intuitions on their internal representations and

---

[30] Note that the terms “bottom-up” and “top-down” used previously map onto Fodor’s (1983) “peripheral” and “central” systems, respectively.

cognitive strategies as they perform complex cognitive activities (cf. the methodology of protocol analysis). The upshot of this is that there is a division of labour in cognitive science: respective researchers "stick to their modules," so to speak, and research on perception often proceeds by trying to factor out the conative aspects of cognition, whereas research on "central" processes often tries to give an account of conation (goals, planning, motivation) with no reference to perception.[31] What both camps have in common is a commitment to elucidating cognitively-mediated behaviour in terms of the internal representations of the world that the cognitive system is supposed to construct.

On the IP paradigm outlined above, attention is seen as but one stage in the visual system, itself supposed to be a module or set of modules in the periphery of the cognitive system. Vision itself is supposed to be entirely stimulus driven until the visual system constructs the primitive visual representations that are fed to the more conceptually-driven "central" systems (Yantis, 1998; Nakayama, He, & Shimijo, 1995). Perceptual organization and attention should not be influenced by conative factors on this approach; at most, conative and other top-down factors will select and modify the output of the peripheral modules, but never influence their internal functioning.

In contrast to the standard, serial IP view of visual attention, I argued above that attention makes use of top-down (conceptual, conative) as well as bottom-up (stimulus-driven) information and knowledge within a unified process, and that this information is

---

[31] Barsalou's Perceptual Symbol System (1999) is an (unfortunately all-too-rare) attempt to integrate perception and higher cognitive processes. Unlike the conative model, however, it is not concerned with giving an account of how conation and perception interact and shape cognition.

not processed serially but rather integrated in a dynamic, multiple soft constraint satisfaction (MSCS) manner. I also argued, *contra* the serial IP approach, that conation (a central process according to the serial modular thesis), conceptual knowledge (another central process) and perceptual organization (a peripheral module) are not individual modules serially chained but rather autonomous but highly interpenetrated systems of constraints that are integrated by visual attention for the purpose of directing visually-guided action. If the conative model of visual attention is correct, the serially modular hypothesis must be abandoned. A new conception of the nature of cognition and of cognitive architecture are required. In the sections that follow I will argue that the model I have proposed for attention – that of a tool-mediated activity with an underlying MSCS-based architecture – can be generalized to the whole of cognition. I will also argue that, in general, the empirical study of cognition must involve the observation of people engaged in meaningful activities.

7.1 Activities and Tools: A New Metaphor for Cognition

Part of the appeal of the standard IP model is that it provides a metaphor for cognition (the computer) that allows us to naturalize human intelligence – intelligent behaviour is not the result of some mysterious Cartesian soul, but rather of the processing of information by computation over representations. It is a metaphor that makes room for both environmental and internal determinants of behaviour, unlike behaviourism, the predecessor to the IP paradigm as the dominant approach in the empirical study of the mind. However, a drawback of the IP approach is that it encourages the study of the internal representations ahead of studying the purposes to which representations are supposedly put. A quick survey of experimental or theoretical work in many areas of cognitive science reveals that representations are studied in light of either highly contrived behaviour (e.g., experiments on visual attention or word recognition) or of *a priori* assumptions about the nature and purposes of human behaviours. To give an example of the latter approach, let us return to Relevance Theory, which was introduced in Chapter 5. On that theory, the interpretation of utterances is guided by the principle that speakers intend their utterances to be relevant to the situation at hand. In developing Relevance Theory, Sperber and Wilson (1995) assume that most people attempt to communicate (i.e., convey some piece of information) when they utter sentences, but provide *no empirical evidence whatsoever for this claim.* Yet, Speech-Act theory, which Relevance Theory itself grows out of, suggests many purposes for speech other than the conveyance of information (Searle, 1969). To paraphrase William James' famous aphorism on attention: everyone knows what cognition is for – and thus we often take it for granted that we only need to study the representational and computational mechanisms that constitute cognition.

There are two important difficulties with focusing on representations before studying the actions they are supposed to produce and the motivations underlying the actions. One is the fact, as Robert Audi reminds us, that "an adequate theory of human behavior must either employ some concept of action and explain its connection with motivation and cognition, or show, as no one has even come close to doing, that our behavior can be understood apart from these crucial notions in action theory" (Audi, 1993, p. 1). Experiments 6 and 7 in Chapter 4 exemplify this point. If the purpose and demands of the experimental task (tracking one set of dots with another by moving the second set of dots with mouse or head movements, respectively) had not been taken into account in the analysis of the data, it would have been difficult to provide a coherent explanation of the results.

Another problem with focusing on representations first is that it introduces into cognition the problems associated with correspondence theories of truth (Audi, 1998), and burdens the study of cognition with solving the basic problems of epistemology. The correspondence theory assumes that truthful representations are isomorphic with reality – what James called the “picture theory” of truth (Hacking, 1983). Similarly, on the standard IP approach, representations enable action by serving as veridical computational surrogates for the world. A focus on representations in the study of cognition thus imposes the burden of accounting for how it is that cognition constructs true – and therefore isomorphic – representations of the world. An isomorphism between perceptual and attentional structures and the world was already discussed and rejected in Chapter 3. As to the difficulties with the correspondence theory of cognitive representations, one need simply to look at the failure of symbolic AI to produce robust and functional robots able to navigate in the world (e.g., Winograd’s *SHRDLU*) or expert systems able to reason about problems beyond narrowly-defined domains of expertise (e.g., Lenat's *CYC*).[32]

A new metaphor is needed which preserves representations as mediators of cognitive activity while, on the one hand, avoiding a correspondence theory of cognitive representations, and on the other, reminding us that representations must be understood in the context of the activities in which they occur and the embodiment of the cognitive agents using them (Clark, 1997). Such a metaphor is provided by the work of a researcher of cognition who was one of the very few working before the advent of the IP paradigm to give an account of the role of representations in cognition (*contra* the behaviourists) that also produced a significant body of empirical work: Lev Semionovich Vygotsky (1896 – 1934).

Vygotsky's work touched on a very wide array of topics, many more than the research on developmental and educational psychology he is usually credited for in the West (Wertsch, 1985; Glick, 1997). Three main strands have been identified in Vygotsky's thinking on cognition: (1) a belief that all cognition must be understood in terms of its development, (2) the claim that "higher" mental processes (what was referred to above as the "central" processes) have their origin in the social nature of human beings, and (3) a conception of cognition as the activity of regulating behaviour with the mediation of tools, both internal and external (Wertsch, 1985). The aspect of Vygotsky's work that concerns us here is the conception of cognition as a tool-mediated activity. For

---

[32] The difficulties encountered by symbolic-AI approaches to real-world problems, as with *SHRDLU* and *CYC*, have been extensively discussed, e.g., in Clark (1997) and Thagard (1996).

Vygotsky, the ‘tools’ included external tools, in their turn including concrete tools and external symbols, and internal tools, i.e., mental representations. On this view, the tools that mediate cognitive activity shape the nature of the activity, but at the same time are shaped by it.

Vygotsky eventually developed a semiotic theory of cognitive function and development, focusing on language and signs as the main tools mediating cognition (Wertch, 1985). His classic example is that of how thinking undergoes a qualitative shift when children start to direct thinking through inner speech, while speech itself becomes more abstract and complex as the child uses it more in thought. On this account, understanding language as a tool mediating thought shows how thought is determined both by social factors (through the influence of language on thought) and by cognitive factors (through the influence of thought on language).

I do not propose to adopt Vygotsky’s later semiotic theory of cognition as an alternative to the IP paradigm, and a discussion of the nature and merits of Vygotsky’s account far exceeds the scope of this work.[33] However, what I wish to borrow from Vygotsky is a general notion of representations as tools mediating cognitive activity. By suggesting that representations enable action by playing the role of tools instead of standing for states of affairs in the world, the tool metaphor avoids the problems associated with the correspondence theory of truth discussed above. Furthermore, since tools are generally shaped by the activities to which they are put, the metaphor helps us

---

[33] Glick (1997), among others, has described the huge exegetical difficulties facing scholars attempting to reconstruct Vygotsky’s theory of cognition.

think of representations in light of the specific behaviours and activities they mediate. Thus the tool metaphor for representations (and its corollary, the tool-mediated-activity metaphor for cognition) fulfills the requirements for a new metaphor[34] for cognition I set out above.

How well does the account of cognition mediated by representations extend beyond Vygotsky's account of the relation between language and thought? In Chapter 5 I suggested that visual attention should be construed as an activity whereby cues to perceptual organization from early vision serve as tools to allocate attention to relevant stimuli. As attention plays a role in perceptual organization, we can see that the tools of attention are shaped by the activity. Furthermore, since the nature of the perceptual tools used by attention shape its behaviour – object-based when object-based perceptual features are selected as relevant, space-based when spatial features are selected as relevant – we also see how the activity is shaped by the tool. The tool metaphor for cognition allows for useful alternatives to the computer metaphor for conceptualizing visual attention and the relationship between language and thought.

The tool-and-activity metaphor of cognition is too vague still to serve as the basis for a general theory of cognition; in particular, the notions of "activity," "tool" and "tool use" have to be further spelled out. However, I propose that the metaphor could be fruitfully generalized to the rest of cognition as a heuristic model for advancing the

---

[34] I use 'metaphor' because the notion of cognition as activities using tools is meant to play the same role as the computer metaphor in guiding research on cognition (a heuristic for generating hypotheses, exploring intuitions, etc.). I recognize, however, that the notion of representations as tools is more an extension of the notion of tools than an outright metaphor.

relevance-based, embodied, action-oriented approach to cognition that is gaining prominence in cognitive science (Clark, 1997; Sperber & Wilson, 1995).

7.2 Interpenetrated Systems of Constraints: A New Cognitive Architecture

The interpenetration of “central” and “peripheral”, or “top-down” and “bottom-up,” information in the conative model is not compatible with a sequentially modular cognitive architecture. Work in other areas of cognition also put pressure on the serially modular view. As mentioned in Chapter 5, in linguistics, Relevance Theory is another account which requires top-down and bottom information to simultaneously contribute to on-line cognitive processing of perceptual input (in this case, spoken language). While still adhering to a modular architecture, Sperber & Wilson have challenged Fodor’s assertion that “central” cognition cannot be explained in terms of information processing (Fodor, 2000) and argue for top-down knowledge playing a computational role in the understanding of utterances.[35] Therefore, what is needed is a cognitive architecture which can account for the interpenetration of processes and information held to be “central” and “peripheral,” and thus segregated, under the standard IP view.

The MSCS architecture proposed as the functional architecture for visual attention in Chapter 5 allows for the interpenetration of top-down and bottom-up processes while maintaining a degree of autonomy of these processes. On the conative model of attention,

---

[35] Whether Sperber and Wilson think that Fodor’s central system can be split into modules, or that processes that on Fodor’s view belong in the central system should be located in the peripheral modules dedicated to language, is unclear. This lack of precision about cognitive architecture in Sperber and Wilson’s account is more circumstantial evidence that the serially modular hypothesis cannot account for meaningful human activity.

perceptual cues, conative factors and conceptual factors each constitute a system of constraints. These systems play the role played by modules in the serially modular theory. Within each system, individual constraints are generally “soft,” i.e., satisfying the set of constraints within a given system does not require satisfying each and every constraint, so long as on balance, the set of constraints is satisfied (Thagard & Verbeurgt, 1998). However, the ability of certain stimuli to elicit responses almost universally (e.g., recognizing one’s given name) suggests that certain constraints are “harder” than others, and must be satisfied by the MSCS mechanism governing the system of constraints (or at the very least, the pattern produced by the MSCS mechanism must be compatible with the constraint). Thus, instead of a hierarchy of processing stages or modules, we have a hierarchy of constraint strengths. Furthermore, this hierarchy is fluid, as the strength of individual constraints would vary in response to activity from other constraint systems and external stimulation (bearing in mind that the harder a constraint is, the less likely it is to vary in strength).[36]

In terms of general cognitive architecture, I propose that particular cognitive functions or activities are realized as MSCS mechanisms which integrate and flexibly satisfy the constraints imposed by the various constraint systems (corresponding to

---

[36] “Hard” constraints in varions cognitive systems might provide a generalized conceptual scheme through which all perception is filtered and which shapes all cognition, much as Kant’s proposed spatiotemporal manifold and the 12 categories structuring all sensible experience (Kant, 1781/1787). How some constraints come to be “hard” or “soft” is beyond the scope of this work, but I speculate that many “hard” constrains are the result of evolutionary pressures, whereas others are the result of a person’s experiences (overlearning, traumatic experiences, etc.). Furthemore, I think it likely that a constraint’s being “hard” or “soft” is a default setting that can be changed over the lifespan of an individual, although “harder” constraints would change much more gradually, if at all.

"central" and "peripheral" modules) required by the activity or function in question. The interpenetration of the constraint systems is guaranteed by the fact that constraint satisfaction between the systems happens at the same time as constraint satisfaction mechanisms are organizing individual constraints in each system. The *inter*-system MSCS mechanism allows partially activated patterns (i.e., partial solutions to constraints) to be passed from one system to another, ensuring that *intra*-system activity is shaped both from without and from within systems, while *inter*-system constraint satisfaction directs the overall behaviour or process. Note that a generalized MSCS cognitive architecture is consistent with accounts that suggest that cortical dynamics in general are governed by non-linear, adaptive mechanisms such as Darwinian neuronal competition (Calvin, 1996), or self-organizing mechanisms (Grossberg, 1994; Kelso, 1994), as such mechanisms can readily be understood in terms of constraint satisfaction mechanisms, and vice-versa.

As an MSCS account of visual attention has already been given, let us examine how this architecture can be applied in other domains. Returning again to the understanding of linguistic utterances, the interaction of background knowledge and semantic/syntactic processes, guided by the relevance principle, posited by Relevance Theory is a good candidate for explanation via an MSCS architecture. Each system of knowledge posited by RT – syntax, semantics, background knowledge – could be realized (or perhaps "encoded") as a separate constraint system. The relevance principle (conceived of as a motivation to communicate effectively) itself could be seen as a strong constraint in the conative system of constraints. The MSCS system connecting the syntactic, semantic, background knowledge and conative systems would ensure that syntax, semantics and general background knowledge are combined in a manner relevant to the context of the conversation to produce understanding of an utterance. Linguistic pragmatics could thus be seen as the result of the activity of the MSCS process connecting the various systems of constraints on language comprehension listed above.

The MSCS architecture is a more natural functional architecture for relevance-based visual attention (as discussed in Chapter 5), and for utterance interpretation as proposed by Relevance Theory, than the serially modular architecture that is often still the mainstay of much work in cognitive science. It is also a more natural fit with the tool metaphor for cognition developed in Section 7.1 than is the serially modular architecture. I therefore propose that the MSCS architecture developed for visual attention in Chapter 5 can be generalized to the rest of cognition, as I previously argued in the case of the tool metaphor.

7.3 Methodological Implications: Studying how People Manipulate Things

Admittedly, there are many points of similarity between the proposal of cognition as tool-mediated activities and the phenomenology movement in philosophy and the embodied or situated cognition movement in more recent years (e.g., Clark, 1997; Hutchins, 1995; Varela, Thompson & Rosch, 1991). However, those lines of enquiry do not have empirical methodologies that explicitly operationalize their main conceptual commitments, or somehow factor them into experimental designs. For instance, Varela and Thompson have either personally carried out or cite research on biological systems in support of their positions, but no one has proposed a way of building “embodiment” or “phenomenology” or “self-organization” into valid experimental research. Similarly, while Clark’s exploration of the relation between embodiment and cognition is an ambitious and thought-provoking synthesis of the work of the German phenomenologists (with particular reference to von Uexküll, also a biologist by training), Vygotsky, and current research in the neurosciences and Artificial Life (especially that of Rodney Brooks), it does not propose a methodological or empirical program of its own.

What is called for is an experimental methodology for studying cognition from a more situated, purposeful, “activity-based” perspective. Much experimental research in cognitive psychology makes use mainly of response times and error rates as dependent measures, as leafing through almost any issue of the *Journal of Experimental Psychology* will reveal. This perusal of the experimental literature will also reveal that the majority of experimental tasks are what could be characterized as epistemic judgement tasks (e.g., comparing features in visual attention experiment, saying whether a string of letters is a word in research on word recognition, stating whether objects displayed on a screen were present in a previously viewed video in memory experiments). Measuring performance in such epistemic judgement tasks is designed to explore cognition in terms of the IP paradigm discussed earlier, but it tells us little about the cognitive processes involved when people meaningfully act upon their environment (Kirsh & Maglio, 1994).

I propose that, in order to truly understand the relation between cognitive processes and action, cognitive phenomena should also be studied with experiments that involve participants actually manipulating elements in their environment. My experiments, and in particular Experiments 6 and 7, were a small step in that direction, and can be used to illustrate the some general principles I propose for experimental research in cognitive science.

To get a clear view of the effects of active engagement with the world on cognition, specific behaviours and abilities should be studied within the context of different activities. Experiments 5, 6 and 7, by comparing the effects of controlling a display and passively observing it on information processing, provide a very simple example of how varying the activity within which a particular behaviour is performed can reveal the effects of action on cognitive processes. Further, much can be gained by observing how performance in a particular activity evolves over time, as the temporal analysis of the results in Experiment 7 (and the failure to do so in Experiment 6) show. As objectively characterizing putative mental constructs is a difficult enterprise at the best of times (Garner, Hake, & Eriksen, 1956; Green, 1992), multiple measures (subjective reports, reaction times, time-series studies, etc.) should be used to study the cognitive processes that are involved in meaningful action upon the environment. Throughout the series of experiments I carried out, I made use of a variety of measures and quantitative techniques – reaction times, accuracy rates, correlations between reaction times and display properties, patterns in error rates and reaction times over time, and the occasional verbal report – in interpreting the behaviour I observed during the course of my experiments. Experiments 4, 6 and 7 in particular, with their complicated patterns of data, showed the value of multiple measures.

The approach I used in Experiments 6 and 7 – observing the effects of allowing participants to control the motion of one of two groups of elements on visual attention – is only one way of studying cognition within the context of meaningful activities in an experimental setting. As studying the effects of manipulation of attention was not an original goal of my research, I did not pursue this line of inquiry any further than Experiments 6 and 7 in Chapter 4. However, visual attention could be further studied in the context of a wide variety of meaningful activities, such as communication using gestures or interacting with a computer interface. As noted above, applied research on visual attention in the context of flight using aircraft simulators has contributed to the understanding of visual attention.[37]

Research on other areas of cognition can also benefit from observing cognition in the context of action. A study by Kirsh and Maglio (1994), for instance, revealed that participants' behaviour while playing Tetris (a very popular video game) cannot be fully understood solely within the context of the standard IP paradigm, and can only be understood by supplementing the IP model with the assumption that some actions are carried out solely for the purpose of facilitating the mental processing required for the game, even if the action does not get the player objectively closer to the goal (aligning various geometric shapes at the bottom of a computer display). The authors called this type of action "epistemic," to distinguish it from the more "pragmatic" actions usually

---

[37] Interestingly, there is some similarity between the approach I am suggesting here for research on *fundamental* aspects of cognition and the user-centred design approach to human-computer interaction promoted by the famous *applied* psychologist and engineer, Don Norman (Norman & Draper, 1986).

discussed in action theory. Thus, studying cognition in the context of playing a video game led to a revision of the standard IP model of cognition and the creation of a new conceptual distinction in action theory.

A final, but crucial, aspect of the methodology I propose is the identification and eventual operationalization of the “tools” used by cognitive processes. I do not have a systematic procedure developed for this; however, I can offer as an example how I decided to think of perceptual grouping as the tools of attention. Clearly, my conceptualization of objects as tools stemmed from my thinking of attention as an activity. But I borrowed the notion of visual object itself from the literature on object perception and visual cognition in general. I am not in a position to give a detailed history of the conceptual evolution of objects in vision research, but I can say that the process started with psychologists of various stripes trying to determine the mental preconditions for the perception of objects in the world. The introspectionists (i.e., Wundt, Titchener, and so on) turned to a systematic observation of phenomenal experience, whereas modern researchers have used a combination of computational considerations (e.g., Biederman, 1995; Marr, 1982, Kellman & Shipley, 1991) and data from neuroscience on the cortical pathways of vision (e.g., the pioneering work of Hubel and Wiesel, 1962). These efforts have been supplemented by psychophysical experiments on the stimulus conditions leading to the experience of perceptual grouping (e.g., Gestalt psychology, or more recently, the work of Irving Rock and Stephen Palmer, among others, summarized in Palmer, 1999).

Two general patterns emerge from the various attempts to characterize visual objects: first, they involve empirical examination of hypotheses about necessary mental preconditions for vision; second, they involve a particular assumption about the function of vision – namely, the veridical perception of objects as they are in the world. Thus, I would suggest that the determination of the “tools” that mediate cognitive activities (what some might call the representations that cognitive processes operate over) involves empirically testing hypotheses about the necessary preconditions of an activity, in light of the putative purposes of the activity in question, through the use of various (hopefully convergent) empirical tools, including behavioural experiments, neuroscience, psychophysics, and computational modeling. To a certain extent, it is difficult to dissociate the study of the activity from the study of the tool and vice-versa, but one can fruitfully use intuitions about one or the other to bootstrap empirical research into both.

8. CONCLUSIONS AND FUTURE DIRECTIONS

8.1 General Summary and Conclusions

The starting premises of the research project I have described above were that attention was object-based, and visual objects were Gestalt groupings. The problem I set out to deal with was how to understand HUDs from an object-based perspective so as to reduce or eliminate cognitive tunnelling through better HUD design, and gain insights into the implications of introducing mixed-FOR HUDs in cockpits. A review of the research literature on object-based attention and on HUD use, however, led me to re-consider Gestalt groupings as selection units for object-based attention. Furthermore, the literature on object-based attention itself contains some problematic assumptions about visual attention, namely, that attention is a stage in visual processing that occurs before top-down knowledge is applied to visual stimuli, but that at the same time attention is supposed to select visual objects which, at least in principle, could require top-down information to be detected. The series of experiments I carried out strengthened these misgivings about object-based attention as typically formulated with each passing experiment.

I carried out seven experiments, which produced the following results: grouping elements by common motion can induce a single-group processing advantage (the so-called “object effect”), either when only one group moves (Experiment 1) or when both groups have separate trajectories (Experiment 5). However, the single-group advantage can be biased towards one or another group (i.e., attention can be asymmetrically distributed) very strongly by a participant's intention to focus attention on one or another group (Experiments 2, 3 and 4).

Focusing attention on a single group had the main effect of inhibiting the processing of information from the parts of the display that were not at the centre of attention (Experiments 3 and 4). It also had the effect of enhancing information processing in attended objects under very specific conditions, namely, when the group was in motion (Experiments 3 and 4), or when the group was static and was the only group with interconnected elements (Experiment 4).

Static grouping factors seemed to have little influence on grouping by motion and attentional focus. However, when participants were required to track the position of one group of elements (moved by a computer program) with a group of elements they controlled themselves, attention was directed onto the group which was more informative or relevant to the task – the group being tracked in the case where participant input was manual (Experiment 6), or the group moved by the participant when head movements were used to control the group (Experiment 7). This last experiment also revealed that the effects of action on attention vary with the manner in which a task is executed, and that the distribution of attention within the same task can vary over time as participants gain experience with the task.

The experimental results I collected during my seven experiments let to a much richer notion of visual attention than the one suggested by the standard object-based attention model. Accordingly, I proposed a conative account of attention, wherein attention is viewed as a purposeful activity directed by an observer's motivations and background knowledge, and is constrained by the task context and the properties of the visual stimuli to which the observer is exposed. In this view of attention, objects are neither Gestalt grouping nor the unconditional basis of attentional selection, but rather they are visual surfaces "inferred" in lawful ways from various input cues by the visual system and then used by the visual system as tools to direct attention to objects and features of the environment that are relevant to an observer in a particular context. Just as man-made tools both shape and are shaped by the tasks for which they are used, "visual objects" both constrain attention, making it look "object-based" under certain conditions, and are shaped by attention. What has been characterized as "object-based attention" in the research literature is but one mode of functioning of attention, a default mode the visual system falls back on in the absence of dominant top-down influences.

The HUD issues that provided an initial impetus to my research – cognitive tunnelling and assessing mixed-FOR HUDs – took on a different complexion in light of the conative approach to attention. Instead of being purely a design-based problem, cognitive tunnelling is more likely the result of task demands and pilot's habitual (usually implicit) strategies for allocating attention, leading a pilot to focus too much attention on particular elements at the expense of attending to other elements which are less relevant to the pilot in light of his or her total context. Tunnelling could therefore be avoided by training pilots to allocate their attention more symmetrically between obviously relevant and not-so-relevant elements as a matter of policy, or by designing HUDs so that pilots are forced (or at least are given some incentive) to attend both to immediately relevant and potentially relevant elements of their visual environment.

In relation to mixed-FOR HUDs, Experiment 7 suggests that head-yoked elements in a mixed-FOR HUD might command too much of a pilot's attention, especially when the pilot is a novice user of mixed-FOR HUDs. However, experience with the HUD, as well as training for attentional strategies, would likely alleviate any tunnelling-like problems that might arise with mixed-FOR HUDs, as is probably also the case with standard aircraft-referenced HUDs.

The results of my experiments, and results from other research, prompted a reconceptualization of attention as a tool-mediated activity. This reconceptualization has a number of implications for the study of cognition in general. From a conceptual point of view, I propose that the view of cognition as primarily an exercise in representation-building, should be replaced with the view that cognition is a collection of interrelated and mutually influencing activities. On this view, the architecture of cognition would be a set of tool-mediated processes (the tools either being external, i.e., concrete objects, internal, such as representations, or social, such as language) that are implemented with, and interact through, multiple soft-constraint satisfaction mechanisms which combine conative and conceptual factors with environmentally-driven constraints. The methodological implications of this view are that the empirical study of cognition must involve the observation of people engaged in the purposeful manipulation of their environment.

8.2 Future Work

Having presented what aim to be comprehensive and well-articulated views of visual attention, on the one hand, and cognition, on the other, much work remains to be done to develop and validate both of these views. First of all, a few of the experiments in Chapter 4 (namely Experiments 6 and 7) need to be replicated with larger numbers of participants to verify that the effects observed in those studies are indeed significant. Secondly, the series of experiments I carried out suggests a number of possible lines of research that could all be fruitfully followed up, regardless of the ultimate validity of my theoretical account. A number of experiments could be designed following the general experimental paradigm I used, as follows:

- the evolution of attentional deployment over time within a task should be studied more extensively, and perhaps become a fundamental feature of most experiments on attention
- the effects of manipulating groups of elements, and in particular the role of control modality (e.g., head vs. hand), and the purpose of the manipulation, must be studied more extensively
- the generalizability of my findings to applied settings (namely HUDs) is yet to be directly confirmed
- features other than colour should be used in the 2AFC task I employed; tasks other than the 2AFC should also be used

In addition, the role of stimulus meaningfulness in attentional allocation needs to be studied. The simple geometric forms I used are useful for exploring certain very basic aspects of attention in the context of action, but, as was noted above more than once, human behaviour tends to be purposeful. Visual attention therefore needs to be studied in the contexts of activities using meaningful stimuli. One obvious way of doing this would be to study attention in the context of HUD use (not just using displays that share certain features with HUDs). HUDs provide a mix of manipulability and meaningfulness that makes them well suited to the study of visual attention, as is attested to by the fact that research on HUDs has revealed interesting things about visual attention, as discussed extensively above.

Finally, the conative theory of attention requires further development. The notion of visual objects as tools for attentional activity must be developed, and the specifics of the conative model of attention itself must be further articulated and tested. Similarly, the notion of a cognitive tool itself must be better articulated, in particular as it relates to internal representations as tools rather than information-bearing or meaning-bearing structures. The notions of organism-driven and conative factors in both attention and cognition need to be clarified. There is much work that can be done in the absence of an articulated account of conation (e.g., studying relevance-based effects in various areas of cognition), but without such an account, we will at best have a collection of conative effects in cognition and no theoretical framework in which to embed them. This should not deter us, however, from studying cognition from a conative point of view, for, as Hacking (1983) points out, even physics, considered by many to be paradigmatic of good science, was chiefly a collection of experimental effects until the great theoretical unifications that began with the Theory of Relativity and Quantum Theory.

REFERENCES

Albert, M. K., & Hoffman, D. D. (1995). Generictiy in spatial vision. In R. D. Luce, M. D'Zmura, D. D. Hoffman, G. J. Iverson, & A. K. Romney (Eds.), *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow on his 70th Birthday*. Mahwah, NJ: Erlbaum.

Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 631-665). Cambridge, MA: MIT Press.

Audi, R. (1993). *Action, Intention, and Reason.* Ithaca, NY: Cornell University Press.

Audi, R. (1998). *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. London & New York: Routledge.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-609.

Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics, 51*, 145-162.

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

Ben-Av, M. B., Sagi, D., & Braun, J. (1992). Visual attention and perceptual grouping. *Perception & Psychophysics, 52,* 277-294.

Bichot, N. P., Cave, K. R., & Pashler, H. (1999). Visual selection mediated by location: Feature-based selection of noncontiguous locations. *Perception & Psychophysics, 61,* 403-423.

Biederman, I. (1995). Visual object recognition. In S. M. Kosslyn and D. N. Osherson (Eds.), *An Invitation to Cognitive Science, Volume 2: Visual Cognition* (2nd ed.). Cambridge, MA: MIT Press.

Broadbent, D. E. (1958). *Perception and Communication*. New York: Pergamon Press.

Broadbent, D. E. (1971). *Decision and Stress.* London: Academic Press.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139-159.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97, 523-547.

Calvin, W. H. (1996). *The Cerebral Code: Thinking a Thought in the Mosaics of the Mind.* Cambridge, MA: MIT Press.

Carpenter, G. A., & Grossberg, S. (1991). *Pattern Recognition by Self-Organizing Neural Networks.* Cambridge, MA: MIT Press.

Castiello, U., & Umiltà, C. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception & Performance, 18,* 837-848.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again.* Cambridge, MA: MIT Press.

Craighero, L., Fadiga, L., Umiltà, C., & Rizzolatti, G. (1999). Action for perception: A motor-visual attentional effect. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1673-1692.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (revised edition). New York: Academic Press.

Cohen, D. J. (1999). Elements or objects? Testing the movement filter hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 348-360.

Cohen, D., Otakeno, S., Previc, F. H., & Ercoline, W. R. (2001). Effect of “inside-out” and “outside-in” attitude displays on off-axis tracking in pilots and non-pilots. *Aviation, Space, and Environmental Medicine, 72,* 170-176.

Crowell, J. A., Banks, M. S., Shenoy, K. V., & Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature, 1,* 732-737.

Dowell, S.R., Foyle, D.C., Hooey, B.L. & Williams, J.L. (2002). The effect of visual location on cognitive tunneling with superimposed HUD symbology. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomic Society,* 121-125. Santa Monica, CA: HFES.

Downing C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner and O. S. M. Marin (Eds.), *Attention & performance: Vol. XI* (171-188). Hillsdale, NJ: Erlbaum.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metphor brakes down. *Journal of Experimental Psychology: Human Perception and performance*, *15*, 448-456.

Driver, J., & Baylis, G.C. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The Attentive Brain.* Cambridge, MA: MIT Press.

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

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

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183-201.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and the functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* Cambridge, UK: Cambridge University Press.

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

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*, 583-597.

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDs: Are they viable? *Human Factors, 43,* 173-193.

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

Fernandez-Duque, D., & Johnson, M. L. (1999). Attention Metaphors: How metaphors guide the congnitive psychology of attention. *Cognitive Science*, *23*, 83-116.

Fischer, E., Haines, R. F., & Price, T. A. (1980). Cognitive issues in head-up displays. *NASA Technical Paper 1711*, NASA Ames Research Center, Moffett Field, CA.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology.*Cambridge, MA: MIT Press.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In Z. Pylyshyn (Ed.), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence.* Norwood, NJ: Ablex.

Fodor, J. A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology.* Cambridge, MA: MIT Press.

Foyle, D. C., Dowell, S. R., & Hooey, B. L. (2001). Cognitive tunneling in head-up display (HUD) superimposed symbology: Effects of information location. In R. S. Jensen, L. Chang, & K. Singleton (Eds.), *Proceedings of the Eleventh International Symposium on Aviation Psychology, 143:1-143:6.* Columbus, Ohio: Ohio State University.

Foyle, D. C., Sanford, B., & McCann, R. S. (1991). Attentional issues in superimposed flight symbology. In R. S, Jensen, (Eds.), *Proceedings of the Sixth*

International Symposium on Aviation Psychology, 577-582. Columbus OH: The Ohio State University.

Garner, W. R., Hake, H. W., & Eriksen, C. W. (1965). Operationism and the concept of perception. *Psychological Review, 63,* 149-158.

Green, C. D. (1992). Of immortal mythological beasts: Operationism in psychology. *Theory & Psychology, 2,* 291-320.

Gibson, B. S., & Egeth, H. (1994). Inhibition of return to object-based and environment-based locations. *Perception & Psychophysics,* 55, 323-339.

Gibson, J. J. (1979). *The Ecological approach to visual perception.* Dallas, TX: Houghton Mifflin.

Girotto, V., Kemmelmeier, M., Sperber, D., & van der Henst, J.-B. (2001). Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition, 81,* B69-B76.

Glick, J. (1997). Prologue. In R.W. Rieber (Ed.), M. J. Hall (transl.), *The Collected Works of L.S. Vygotsky, Volume 4: The History of the Development of Higher Mental Functions.* New York: Plenum Press.

Goldsmith, M. (1998). What’s in a location? Comparing object-based and space-based models of feature integration in visual search. *Journal of Experimental Psychology: General*, *127*, 189-219.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action.*Trends in Neurosciences, 15,* 20-25.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. In *Perception & Psychophysics, 55,* 48-120.

Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: interactions of boundary, surface, spatial and object representations. In *Psychological Review, 101*, 470-489.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science.* New York: Cambridge University Press.

Harman, G. (1965). The inference to the best explanation. *The Philosophical Review, 74,* 88-95.

Hammond, M., Howarth, J., & Keat, R. (1991). *Understanding Phenomenology*. Oxford, UK: Basil Blackwell Ltd.

Hardcastle, V. G. (2003). Attention versus consciousness: A distinction with a difference. In N. Osaka (Ed.), *Neural Basis of Consciousness* (pp. 105-120). Amsterdam: John Benjamins Publishing Company.

Heelan, P. A. (1983). *Space-Perception and the Philosophy of Science.* Berkeley: University of California Press.

Hempel, C. G. (1966). *Philosophy of Natural Science.* Englewood Cliffs, N.J.: Prentice-Hall.

Herdman, C. M., Jarmasz, J., and Johannsdottir, K. (2000). *Research on Heads Up Displays and Helmet Mounted Symbology.* Summary Report. PWGSC File Number: WW7711-9-7577.

Herdman, C. M., Johannsdottir, K. R., Armstrong, J., Jarmasz, J., LeFevre, J., & Lichacz, F. (2001a). Mixed-up but flyable: HMDs with aircraft- and head-referenced symbology. *Engineering Psychology and Cognitive Ergonomics: Ashgate.*

Herdman, C. M., Johannsdottir, K., Lessard, L., Jarmasz, J., Churchill, L.L., Farrell, F. R. (2001b). Attentional benefits and costs associated with integrating a direct voice input (DVI) system into a multi-crew helicopter environment. *Proceedings of the Eleventh International Symposium on Aviation Psychology.*

Hoffman, D. D. (1998). *Visual Intelligence.* New York, NY: W. W. Norton & Company.

Holyoak, K. J., & Thagard, P. (1996). *Mental Leaps: Analogy in Creative Thought.* Cambridge, MA: MIT Press.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24,* 849-878.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology of London, 160*, 106-154.

Humphreys, G. W., (1993). Interaction between objects and space systems revealed through neuropsychology. In D. E. Meyer & S. Kornblum (Eds.), *Attention & performance XIV* (143-162). Cambridge, MA: MIT Press.

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, MA: MIT Press.

Jarmasz, J., Herdman, C. M., & Johannsdottir, K. R. (2001). Object layers in HUDs: The role of motion in grouping symbology. *Engineering Psychology and Cognitive Ergonomics: Ashgate.*

Johnson, D. N., & Yantis, S. (1995). Allocating visual attention: Tests of a two-process model. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1376-1390.

Jonides, J. P. (1981). Voluntary versus automatic control over the mind's eye. In J. Long and A. Baddeley (Eds.), *Attention & Performance IX.* Hillsdale, NJ: Erlbaum.

Jordan, H., & Tipper, S. P. (1998). Object-based inhibition of return in static displays. *Psychonomic Bulletin & Review, 5*, 503-509.

Julesz, B. (1990). Early vision is bottom-up, except for focal attention. *Cold Spring Harbor Symposia on Quantitative Biology – The Brain, 55,* 973-978.

Julesz, B. (1991). Early vision and focal attention. *Reviews of Modern Physics, 63*, 735-772.

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., & Treisman, A. M. (1984). Changing view of attention and automaticity. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), *Varieties of Attention*. New York: Academic Press.

Kahneman, D., Treisman, A. M., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175-219.

Kant, I. (1781/1787). *Critique of Pure Reason.* Trans. N. K. Smith as *Immanuel Kant's Critique of Pure Reason*. London: MacMillan, 1963.

Kanizsa, G. (1979). *Organization in Vision; Essays on Gestalt Perception.* New York: Praeger.

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

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. In *Cognitive Psychology, 23,* 141-221.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18,* 513-549.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate.* Cambridge, MA: MIT Press.

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt, Brace & World.

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

Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10,* 1-40.

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

Levy, J. L., Foyle, D. C., & McCann, R. S., (1998). Performance benefits with scene-linked HUD symbology: An attentional phenomenon? *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society, 11-15.* Santa Monica, CA: HFES.

Leyton, M. (1992). *Symmetry, Causality, Mind.*Cambridge, MA: MIT Press.

Lipton, P. (1991). *Inference to the Best Explanation.* London: Routledge.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476-490.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. Psychological Review, 103, 603-649.

Loux, M. J. (1998). *Metaphysics: A Contemporary Introduction.* London: Routledge.

Lowe, D. G. (1985). *Perceptual Organization and Visual Recognition.* Boston, MA: Kluwer Academic Publishers.

Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992), Perceptual organization and attention. *Cognitive Psychology*, *24,* 475-501.

Mapelli, D., Cherubini, P., & Umiltà, C. (2002). Attending to objects: Costs or benefits? *Acta Psychologica, 109,* 57-74.

Marr, D. (1982). *Vision; a Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: Freeman.

Martin-Emerson, R., & Wickens, C. D. (1997). Superimposion, symbology, visual attention, and the head-up display. *Human Factors*, *39*, 581-601.

McCann, R. S., & Foyle, D. C. (1994). Superimposed symbology: Attentional problems and design solutions. *SAE Transactions: Journal of Aerospace, 103,* 2009-2016.

McCann, R. S., & Foyle, D. C. (1995). Scene-linked symbology to improve situation awareness. *AGARD Conference Proceedings CP 575* (16-1 - 16-11). Brussels, Belgium.

McCann , R. S., Foyle, D. C., & Johnston, J. C. (1993). Attentional limitations with head-up displays. In R. S. Jensen (Ed.), *Proceedings of the 37th Annual Meeting of the Human Factosr Society* (1345-1349). Columbus, OH: The Ohio State University.

Milner, A. D., & Goodale, M. A. (1998). Precis of *The Visual Brain in Action*. *Psyche, 4(12),* available online at http://psyche.cs.monash.edu.au/v4/psyche-4-12-milner.html.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. In S. M. Kosslyn and R. A. Andersen (Eds.), *Frontiers in Cognitive Neuroscience*, p. 19-23. Cambridge, MA: MIT Press, 1992.

Moray, N. (1959). Attention and dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11,* 56-60.

Myors, B. (1999). Timing accuracy of PC programs under DOS and Windows. *Behavior Research Methods, Instruments, & Computers, 31*, 322-328.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. N.

Osherson (Eds.), *An Invitation to Cognitive Science, Volume 2: Visual Cognition* (2nd ed.). Cambridge, MA: MIT Press.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Norman, D. A., & Draper, S. W. (1986). *User Centered System Design: New Perspectives on Human-computer Interaction.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Palmer, S. E., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1,* 29-55.

Peirce, C. S. (1878). How to make our ideas clear. *Popular Science Monthly, 12,* 286-302.

Peirce, C. S. (1903/1997). *Pragmatism as a Principle and Method of Right Thinking: The 1903 Harvard Lectures on Pragmatism.* Edited by P. A. Turrisi. Albany, NY: State University of New York Press.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization. In, K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of Perception and Human Performance, Volume II*. New York, NY: John Wiley & Sons, Inc.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention & Performance, Volume 10* (531-554). Hillsdale, NJ: Erlbaum.

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

Pylyshyn, Z. (1998). Visual indexes in spatial vision and imagery. In R. D. Wright (Eds.), *Visual Attention* (215-231). Oxford, NY: Oxford University Press.

Pylyshyn, Z. (2000). Visual indexes, preconceptual objects, and situated vision. Manuscript under review for a special issue of *Cognition* on "Objects and Attention," July 2000.

Pylyshyn, Z., Burkell, J., Fisher, B., Sears, C., Schmidt, W., & Trick, L. (1994). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology, 1994, 48*, 260-283.

Pylyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 1-19.

Ravden, S. J., & Johnson, G. I. (1989). *Evaluating Usability of Human-computer Interfaces: A Practical Method.* Chichester, UK: Ellis Horwood.

Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: Results of a new method. *Cognitive Psychology*, *24*, 502-534.

Sarter, N. B., & Woods, D. D. (1991). Situation Awareness: A Critical but Ill-Defined Phenomenon. *The International Journal of Aviation Psychology, 1*, 45-57.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language.* London: Cambridge University Press.

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

Selman, B., Levesque, H. J. (1996). Support set selection for abductive and default reasoning. *Artificial Intelligence, 82,* 259-272.

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

Shannon, C. E. (1938). “A symbolic analysis of relay and switching circuits.” Master’s thesis, Massachusetts Institute of Technology; published in *Transactions of the American Institute of Electrical Engineers, 57*, 1-11.

Shelden, S. G., Foyle, D. C., & McCann, R. S. (1997). Effects of scene-linked symbology on flight performance. Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society, 294-298. Santa Monica, CA: HFES.

Sheliga, B. M., Riggio, L., & Rizzolati, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research, 98,* 507–522.

Smith, B. C. (1996). *On the Origin of Objects.* Cambridge, MA: MIT Press.

Solso, R. L. (1994). *Cognition and the Visual Arts.* Cambridge, MA: MIT Press.

Spearman, C. (1937). *Psychology Down the Ages.* London: Macmillan.

Spelke, E. S., Gutheil, G., & Van de Walle, G. (1995). The development of object perception. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual Cognition: An Invitation to Cognitive Science Vol 2* (297-330). Cambridge, MA: MIT Press.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.

Sperber, D., & Wilson, D. (1996). Fodor's Frame Problem and Relevance Theory (reply to Chiappe & Kukla). *Behavioral and Brain Sciences, 19,* 530-532.

Stelmach, L. B., Campsall, J. M., & Herdman, C. M. (1997). Attentional and Occular Movements. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 823-844.

Sternberg, S. (1969). Memory-scanning : Mental processes revealed by reaction-time experiments. In *American Scientist, 57,* 421-457.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 28,* 643-662.

Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge, MA: MIT Press.

Thagard, P. (1996). *Mind: Introduction to Cognitive Science.* Cambridge, MA: MIT Press.

Thagard, P. & Shelley, C. (1997) Abductive reasoning: Logic, visual thinking, and coherence. In M.-L. Dalla Chiara et al. (Eds), *Logic and Scientific Method* (p. 413-427). Dordrecht: Kluwer.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science, 22*, 1-24.

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

Tipper, S. P., & Weaver, B. (1998). The medium of attention: Location-based, object-based, or scene-based? In R. D. Wright (Ed.), *Visual Attention* (77-107). Oxford, UK: Oxford University Press.

Treisman, A. (1988). Features and objects: The fourteenth Barlett memorial lecture. *Quarterly Journal of Experimental Psychology, 40A,* 201-237.

Treisman, A. (1998). Feature binding, attention and object perception. In *Philosophical Transactions of The Royal Society of London, 353,* 1295-1306.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136.

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

Tsal, Y. (1994). Effects of attention on perception of features and figural organisation. *Perception, 23,* 441-452.

Tsotsos, J.K., Culhane, S., Wai, W., Lai, Y., Davis, N., Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78,* 507-547.

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

Vecera, S. P. (2000). Toward a Biased Competition Account of Object-Based Segregation and Attention. *Brain and Mind, 1,* 353–384.

Vecera, S. P., Vogel, E. K., & Woodman, G. F. (2002). Lower region: A new cue for figure-ground assignment. *Journal of Experimental Psychology: General, 131,* 194-205.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: MIT Press.

Wallis, G. (2002). The role of object motion in forging long-term representations of objects. *Visual Cognition, 9,* 233-247.

Wertsch, J. V. (1985). *Vygotsky and the Social Formation of Mind.* Cambridge, MA: Harvard University Press.

Wickens, C. D., & Long, J. (1994). Conformal symbology, attention shifts, and the head-up display. *Proceedings of the 38th Annual Meeting of the Human Factor and Ergonomics Society,* (6-10) Nashville, TN: Human Factors and Ergonomics Society.

Wickens, C. D., & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of Head-Up Displays. *Journal of Experimental Psychology: Applied, 1,* 179-193.

Wright, R. D., & Ward, L. M. (1998). The control of visual attention. In R. D. Wright (Ed.), *Vancouver Series in Cognitive Science, vol. 8: Visual Attention* (132-186). Oxford, NY: Oxford University Press.

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

Yantis, S. (1998). Objects, attention, and perceptual experience. In R. D. Wright (Ed.), *Vancouver Series in Cognitive Science, vol. 8: Visual Attention*, 187-214. Oxford, NY: Oxford University Press.

Zemel, R., Behrmann, M., Mozer, M. C., & Bavelier, D. (2002). Experience-Dependent Perceptual Grouping and Object-Based Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 202-217.