Guiding attention in a dynamic environment

Pinto, Y.

2008

citation for published version (APA)
Pinto, Y. (2008). *Guiding attention in a dynamic environment*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

“A picture says more than a thousand words”. This proverbial saying is especially true when it comes to memory. As anyone familiar with computers knows, storing a picture really takes up the equivalent space of an entire book. Therefore, if humans would retain every image they have seen, the brain would soon be flooded with information. To retain what is important, and discard what is not, a selection has to be made. This selection procedure corresponds to what we subjectively experience as attention. In general, we can consciously remember only what we attend to (Becker, Pashler, & Anstis, 2000; Landman, Spekreijse, & Lamme, 2003). Thus, attention can be viewed as the gatekeeper. Information that does not get passed this gate essentially disappears into oblivion.

As an important cognitive function, attention has been extensively studied since the 1950’s (for an overview see Pashler, 1998, or Driver, 2001). A key question in attention research is which factors drive selection. What makes you attend to one object and not the other? Theories essentially distinguish between two types of factors: factors relating to the observer, and factors relating to the stimulus. Factors generated by the observer are referred to as top-down factors. Goals of the observer (such as looking for rectangular objects) are an example of a top-down factor. The color, the shape, or the speed of an object are stimulus properties, and these are referred to as bottom-up factors. For example, imagine you are looking for your friend in a busy train station. You know he is wearing a red sweater, and therefore you decide to look for anything red. In that case, if you attend to a red item, this is because of your search goal. This then would be top-down driven attention. Conversely, in the same situation, a loud bang would catch your attention, even though you did not look for loud bangs – thus providing an example of bottom-up driven attention.

Is attention essentially top-down or bottom-up driven? Three general theoretical frameworks have emerged: The relative bottom-up model, the absolute bottom-up model and the top-down model. We will first outline what these theories encompass, and what empirical evidence supports them. We will then review two different visual settings, involving dynamic displays and dynamic changes between displays, and how well these theories fit the data found in these situations.

Visual search

In all cases, an important (though of course not the only) measure of attention is performance in a visual search task, and the search slopes it produces. In a visual search task, a participant looks for a target item among distractor items. The search slope is the increase in reaction time as a function of the number of items in the display. The search slope obtained in visual search depends on the task. If participants can immediately distinguish the target from the distractors, the number of distractors does not affect reaction times. However, if participants cannot attend the target immediately, they will have to shift their focus from item to item, causing reaction times to be longer if there are more items. For instance, when participants search for a square among circles, the number of non-target elements (circles) does not affect reaction times. Therefore, in this case the search slope is flat, and search is considered to be efficient. Note that in this example, the square is also referred to as a singleton (Pashler, 1988). Singletons denote any item that is unique along a certain dimension (in this case, the square is unique with respect to the shape dimension). Conversely, when participants search for a 2 among 5s (i.e. alarm clock numbers), the more 5s there are, the longer it takes. Thus, such a task produces a positive search slope, and is referred to as inefficient search. Treisman and Gelade (1980) based their influential Feature Integration Theory (FIT) on the striking finding that only objects that carry a unique feature yield efficient search. FIT argues that this distinction occurs, because attention is necessary to bind features. Therefore, if the target does not carry a unique feature, each object in the display has to be attended to determine which *conjunction* of features makes up the target (e.g. 2s and 5s are conjunctions of the same line segments). Although Treisman and Gelade’s theory is not undisputed, the important point is that discovering which search tasks yield efficient search can be revealing about the role of attention in visual processes.

In addition to search efficiencies, search asymmetries are also an important tool for studying the role of attention. Search asymmetries are said to occur when the search for stimulus A among stimuli B is more efficient than the reverse case. For instance, it is easier to find an orange object among red objects than vice versa. Similarly, a Q among O’s is more efficiently found than an O among Q’s. Treisman and Gormican (1988) hypothesized that search asymmetries reveal the presence or absence of a basic feature.

Thus, a Q can be easily found among O's because the target has an added feature. In the reverse case, the O is lacking a basic feature, and since it is easier to detect the presence of a feature than its absence, the O is more difficult to find. Similarly, according to this notion, the orange target can be detected by searching for added "yellowness". Again this theory is not undisputed, but the main point is that in addition to search efficiencies, search asymmetries can also be telling about how attention influences perception. Particularly, search asymmetries seem to show that it is easier to detect the presence than the absence of a feature (see: Wolfe, 2001 for an overview of research on search asymmetries). As such, search asymmetries are a tool to investigate what constitutes a feature. For instance, Kristjánsson and Tse (2001) argued that "bumpiness" is a basic features based on search asymmetries.

Relative bottom-up model

According to the relative bottom-up model, attention is primarily driven by stimulus properties. However, according to this model stimuli in isolation do not determine where attention is allocated, but stimuli in comparison to surrounding stimuli determine where attention goes. Particularly, in the spatial domain, stimuli that *differ* sufficiently on a dimension from other stimuli (i.e. contain a unique feature relative to the background, whereby feature-contrast has to be local) attract attention (Nothdurft, 1993; Itti & Koch, 2000, 2001). So a red object surrounded by other red objects does not draw attention, but this same object surrounded by green objects does. One important line of research supporting the relative bottom up model is the work of Theeuwes (1991; 1992). In his experiments participants searched for a diamond among circles (or vice versa). All elements were green, but on some trials one of the non-target objects (i.e. a circle when participants searched for a diamond) was red (see Figure 1.1). Participants were aware that the red item was never the target, and that they should ignore this potentially distracting item. Nonetheless, when a red distractor was present, search times to find the target increased, indicating that participants could not ignore the red item. This phenomenon is referred to as *attentional capture*. Thus, if certain items attract attention, even though you do not want to attend to these items, attention is said to be captured. The Theeuwes' studies provide an example of this attentional capture, since attention goes to an item that is never the target, and thus that participants (presumably) do not want to attend (but see Bacon & Egeth, 1994). Another example of attentional capture are the studies of Yantis and Jonides (1984), and Jonides and Yantis (1988), which we will discuss more elaborately later on. Here it suffices to state that in these studies attention also goes to an object that is hardly ever the target (and thus that participants do not want to attend).

In contrast to the spatial domain, in the temporal domain stimuli that are the *same* as previous stimuli appear to draw attention. Research supporting the claim that temporal comparisons determine where attention goes was performed by Maljkovic and Nakayama (1994; 1995; 1996). In their study participants searched for an object with a unique color, e.g. one red diamond among several green diamonds or vice versa. When the target color on the current trial was the same as the color of the target on previous trials, participants responded faster. This occurred even when the current target color was the same as the target color seven trials back, although the conscious recollection did not go further back than one trial. Furthermore, when the target color switched every second trial, and participants were made aware of this, participants were still faster on repetition trials than on switch trials, even though they knew that a switch would come. Maljkovic and Nakayama's studies indicate that if a stimulus is the same as a previously observed stimulus, it attracts attention, irrespective of the search goals of the participant.

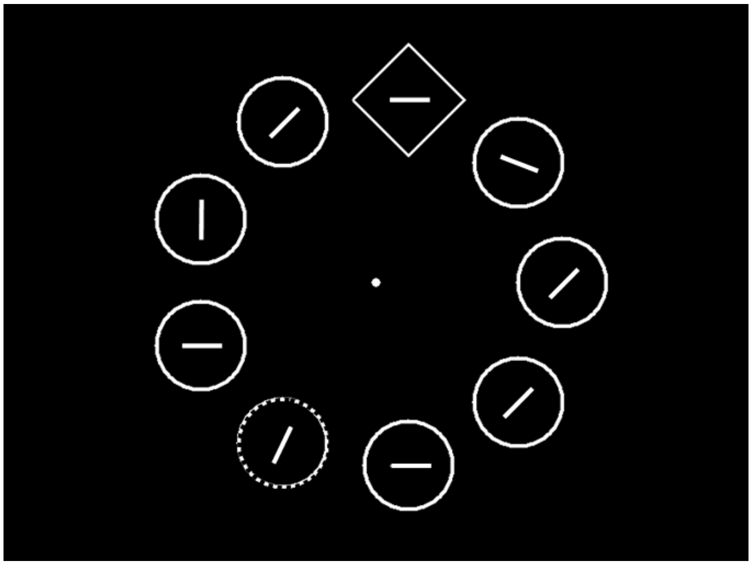

Figure 1.1. Example of a display used by Theeuwes (1992). In his experiment the figures were green and the background was white. The dotted line indicates a red object. In this example participants searched for a diamond (and reported the orientation of the line inside it), while a distractor (the red circle) was present.

Absolute bottom-up model

The absolute bottom up model agrees with the relative bottom up model in the sense that attention is essentially driven by stimulus properties. However, the absolute bottom up model suggests that specific *types* of unique stimuli summon attention not the fact that one element is a singleton (i.e. that the element is different from all elements in the display). According to the prevalent versions of these theories, the most important 'attention-grabbing' stimuli are dynamic objects. Supposedly, from an evolutionary viewpoint, dynamic stimuli indicate potential danger, such as the appearance of a predator, or the sudden looming of an obstacle on our path. Therefore we are inclined to attend to these stimuli. Classic evidence for this model comes from work done by Yantis and Jonides (1984) and Jonides and Yantis (1988). In their studies participants first viewed a display only consisting of figure eights. After some time these figure eights turned into letters. Participants then searched for a certain letter among distractor letters (e.g. an H among P, E and S, see Figure 1.2). The number of distractor letters varied from trial to trial. The more distractor letters were present, the longer it took participants to find the target, indicating that a letter had to be attended before participants knew whether it was the target or not. On some trials, one of the letters was presented with a sudden onset. Most of the times, this sudden onset was not the target, and therefore participants had no incentive to attend to this object. Nevertheless, when the sudden onset happened to be the target, reaction times became independent of the number of distractor letters, indicating that the sudden onset was always attended first. Importantly, this result was only found for sudden onsets. For instance, in the same set-up, when the target sometimes had a unique color, participants were no more efficient in finding it. Yantis and Jonides' research suggests that not all properties attract attention. Onsets do, but unique colors do not. In a similar line of research, Franconeri and Simons (2003) suggested that other dynamic properties such as looming and motion also capture attention. They hypothesized that features capture attention as long as they denote something ecologically relevant.

In this context it is important to mention the study of von Mühlenen, Rempel, and Enns (2005). According to von Mühlenen et al., the key factor is not the dynamic features that are involved, but the *timing* of the dynamics. For instance, they showed that stimuli that captured attention according to Franconeri and Simons, only did so when they changed when everything else was static. That is, when an object loomed at the moment that the search display appeared (i.e. the figure eights turned into letters) then the looming object did not capture attention. However, when it loomed before the search display appeared (as was the case in the study of Franconeri and Simons), then it did capture attention. Similar results were found for objects with a unique color. When an object acquired a unique color before or after the appearance of the search display (i.e. when it was the only changing object) it captured attention. However, when the object acquired the unique color during onset of the search display it did not capture attention. This led von Mühlenen et al. to argue that temporal factors (in addition to spatial factors) are essential for attentional capture. The most effective stimulus is unique in both dimensions, that is, it undergoes a change during relative calm. Von Mühlenen et al.'s notion could be considered to support the relative bottom-up model, since it suggests that comparisons (in both the temporal and spatial domain) are essential for attentional capture.

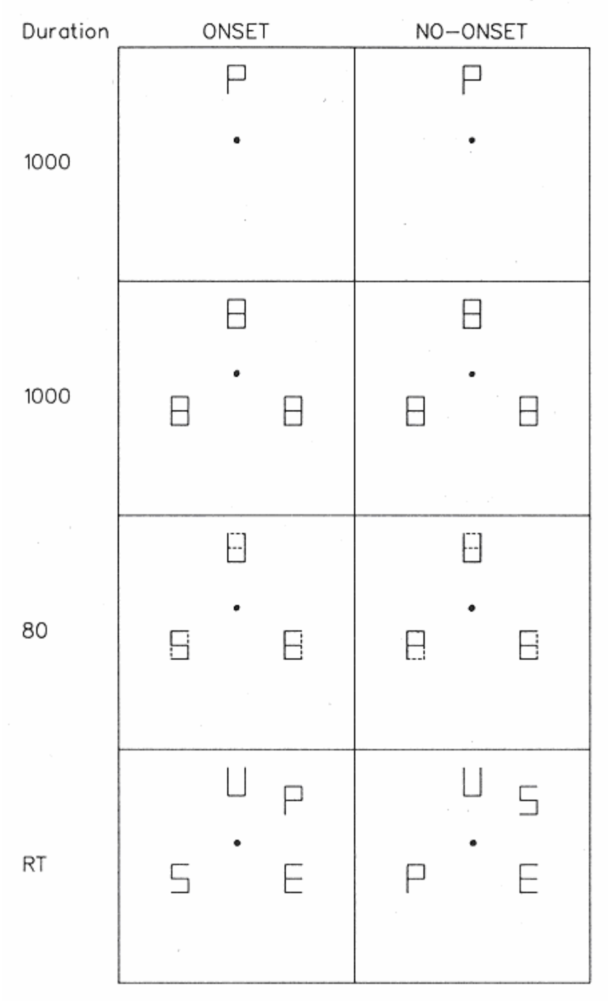

Figure 1.2. An example display as used by Yantis and Jonides (1984), taken from their manuscript. In this example participants reported the presence of the letter P, which could appear as an onset, or on the place previously occupied by a figure eight.

Top-down model

The third influential model is the top-down model. This view states that attention is primarily driven by factors relating to the observer, such as search goals and internal preferences. Again, there are several studies supporting this theory. Folk, Remington and Johnston (1992) presented participants with a search display in which they either had to find a target that appeared with a sudden onset, or a target that carried a unique color. The target could appear in one out of four locations. Before the search display was presented, a cue was shown on one of these locations. The cue appeared 150 ms before the search display, and could either be a sudden onset or a uniquely colored item (see Figure 1.3). This cue was not predictive of where the target would appear, and participants were made aware of this (and were thus instructed to ignore the cue). Folk et al. found that the cue only affected reaction times when it was from the same category (onset or unique color) as the target. That is, when participants searched for an onset, a uniquely colored cue could appear anywhere without affecting reaction times. However, an onset cue speeded responses if it appeared on the target location, and slowed responses when it appeared somewhere else. The same pattern held when participants searched for a uniquely colored target (in that case uniquely colored cues did, but onset cues did not affect their reaction times). Folk et al. concluded that features only attract attention if they match the search goals of the subject.

Bacon and Egeth's (1994) study corroborated this hypothesis. They suggested that in Theeuwes' earlier-mentioned studies participants were distracted because the distractor essentially met their search goals. According to Bacon and Egeth, since participants searched for a unique form (i.e. a diamond is unique among circles), they may have searched for anything unique, rather than for diamonds specifically. Since a red object among non-red items is also unique, this then caused them to attend to this distractor. To avoid that participants adopted such a search strategy, Bacon and Egeth construed displays in which several targets were present. For instance, participants indicated the orientation of a line inside a diamond, and then were presented with a display containing three diamonds (and several circles, see Figure 1.4). In such a search display, there are no unique objects, and therefore unicity cannot be a search goal. Bacon and Egeth varied the number of targets from trial to trial, so that on a small portion of the trials there was only one target. They found that under these conditions, when one of the distractors carried a unique color, it did not affect search times, even when there was only one target. Bacon and Egeth took this as evidence that what was previously considered to be proof of bottom-up capture of attention, was in fact another example of top-down driven attention (but see Theeuwes, 2004). Another important study taken to support the theory of top-down control of attention was performed by Yantis and Egeth (1999). In their study participants performed a difficult search task, for instance involving a vertical bar target among tilted bar distractors. Search was aided when the target carried a unique color, but only if color and target coincided frequently. That is, color information was only used, when on most trials the item with the unique color was the target. This suggests that attending to a salient feature, such as a unique color, is under top-down control. People only do this when they see it as beneficial. Thus, knowledge about the target influences where attention goes, again indicating that top-down settings (in this case search knowledge rather than search goals) determine the allocation of attention.

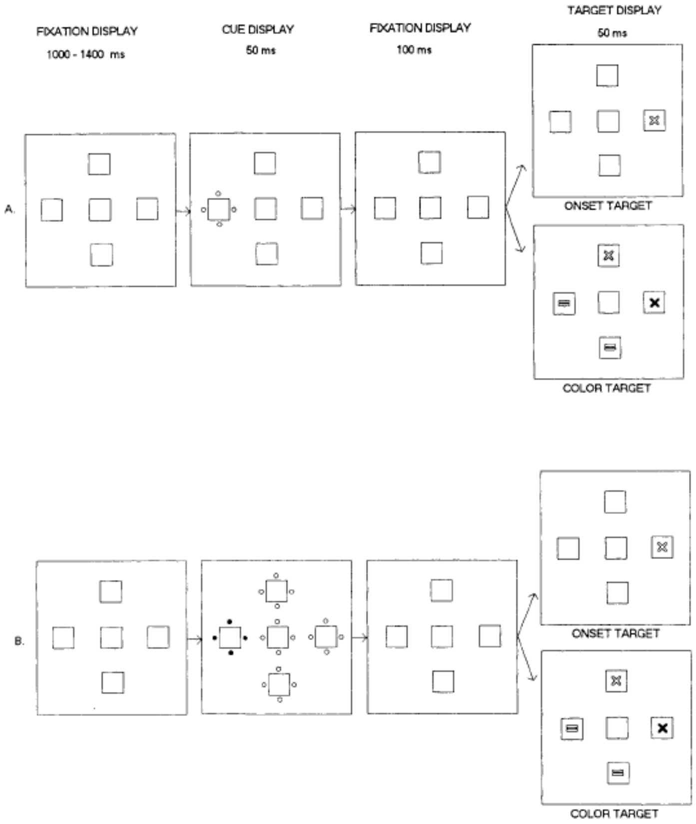

Figure 1.3. Illustration of the study performed by Folk, Remington and Johnston (1992), taken from their manuscript. The cue appeared 150 ms before the search display and could either be an onset (top) or uniquely colored (bottom).

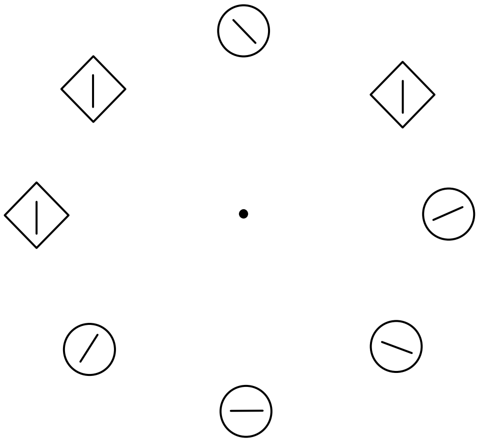

Figure 1.4. Example display of a multi-target display, employed by Bacon and Egeth (1994). In this case participants reported the orientation of the line inside the diamond target.

Rationale behind the present thesis

Summarizing, the three outlined models are not consistent regarding which factors determine attentional allocation. In the following I will review how well these models hold up in conditions involving dynamic display changes within and between trials. The focus is on dynamic displays, because if there has been an absolute attention-puller proposed, it is stimulus dynamics. Dynamic displays will therefore provide the ultimate arena for testing relative versus absolute accounts of attentional capture. First, I will pitch the relative bottom up model against the absolute bottom up model, when the displays are dynamic, and the target is static. Second, I will compare the bottom-up models to the top-down model, in a situation in which the display changes from trial to trial. Such changes create bottom-up effects that may explain phenomena previously attributed to static, top-down settings applied to an entire block of trials.

Attention in dynamic situations

Dynamic among static

Everyday experience suggests that dynamic objects are attention attractors. That is why we wave when we are in a crowded place and we see a friend is looking for us. Similarly, ambulances are equipped with flashing lights, because these are thought to capture attention so strongly that we cannot miss them. According to the absolute bottom up model, this intuition is correct. This model states that dynamic properties are the prime example of attention-attractors. As argued earlier, dynamic objects are thought to attract attention, because they indicate potential danger. A moving or suddenly appearing object can be a predator. Conversely, for a predator, any motion may reveal something about the prey (see Abrams & Christ, 2003; Hillstrom & Yantis, 1994; Tipper & Weaver, 1998, for similar arguments). Previous research has provided confirming evidence for the notion that dynamic items capture attention. Using a visual search task in which observers had to detect a moving target among static distractors, Hillstrom and Yantis (1994) and Yantis and Egeth (1999) found no effect of the number of distractors on search reaction times (RTs; i.e. search slopes were flat), indicating that the moving target could be found efficiently and in parallel across the displays used. Similarly, Mcleod, Driver and Crisp (1988) showed that when there is more than one moving item in the display, search is confined to the moving set and the static set is effectively ignored. Search was also efficient when the type of dynamics was not motion but looming. Hillstrom and Yantis (1994) found that a looming target among static distractors is efficiently found. Interestingly, the same authors found that a third type of dynamics - a “scintillating” noise pattern - was also efficiently found.

Another important type of dynamics is the abrupt onset of an object. We already mentioned the studies of Yantis and Jonides, who found efficient search for an abrupt onset even when it was not the target. Not surprisingly then, Watson and Humphreys (1995) found very efficient search when the target was always defined by an abrupt onset relative to a set of static distractors. Similarly, Watson and Humphreys (1997) later found that search can be restricted to a whole set of new onsets among a set of (at least by then) old static distractors, although it deserves mentioning that Watson and Humphreys explained this effect in terms of inhibition of the non-onsets. Donk and Theeuwes (2003) found that participants attended the new onsets before they started attending the static items, even when there were up to fourteen objects abruptly appearing. These studies provide evidence that attention can be immediately directed towards onsets.

The physiology of dynamics

We might understand the attractiveness of dynamic items better, if we consider how, at a physiological level, the brain treats dynamic objects. Already at a computational level, the processing of dynamic stimuli is different from the processing of static stimuli. When the visual system is confronted with a static object, this initially causes a cascade of neural activity corresponding to the features of that object (e.g. Livingstone & Hubel, 1988). However, after some time the responding neurons fatigue, and neural response to the static object declines. Neural response to a dynamic stimulus is somewhat different. Initially, the dynamic object triggers the same neural response as a static object. But due to its change, it continues to excite different neurons. Hence the neural response to a dynamic stimulus suffers less from fatigue than neural response to static input. Moreover, to detect motion, static feature analysis does not suffice. The mere activation of a retinal cell by itself is not enough to discern whether an object is moving, let alone in which direction. This problem can only be solved if early in the process, the input of several cells is compared. For instance, by comparing changes in luminance at one point in the retina, and correlating this to changes in luminance at neighboring point in the retina after a delay, the visual system could detect a moving stimulus and its direction (this mechanism is in fact carried out by so-called Reichardt cells, see for instance van Santen & Sperling, 1985). The computational differences are reflected physiologically, as there are two major subsystems in the brain preferentially dealing with either dynamic or static stimuli: the processing of dynamic aspects of stimuli takes mainly place in the magnocellular system, whereas static, or more sustained, properties are preferentially analyzed in the parvocellular system (e.g. Breitmeyer & Ganz, 1976; Livingstone & Hubel, 1988; Zeki, Watson, Lueck, Friston, Kennard & Frackowiak, 1991). This diversion starts already at the retina. Light that enters the eye excites the photoreceptors at the retina, and this electric current is subsequently transmitted to ganglion cells located just behind the retina. There are two types of ganglion cells. M (for *magni* or large) cells that detect primarily whether an object is dynamic or not, and P (for *parvi* or small) cells that detect the wavelength (i.e. the color) and the form of an object. The ganglion cells send their signals to the Lateral Geniculate Nucleus (LGN). The LGN is retinotopically organized (i.e. each retinal cell has a corresponding cell in LGN). Moreover, the LGN is divided up in M- and P-layers, which react to either M-cells or P-cells. Thus, in the LGN there is a strict division between the processing of dynamic and static visual information. This separation remains intact in the primary visual cortex (V1), since different layers in the LGN project to different layers in V1. This strict division between the processing of dynamic and static information is referred to as the M and P pathways. After V1 the P-pathway mainly continues in the ventral cortical pathway that extend to the inferior temporal cortex (also called the what-pathway), whereas the M-pathway becomes the dorsal pathway that extends to the posterior parietal cortex (called the how-pathway). Until the primary visual cortex (V1) the most important dynamic stimuli, onsets and moving objects, are treated alike (Andersen, 1997). This common physiological basis is reflected psychophysically. Adaptation to moving stimuli has been shown to impair the perception of blinking stimuli, and vice versa (Chapman, Hoag, & Giaschi, 2004; Green, 1981). Furthermore, at an attentional level, Folk, Remington and Wright (1994) found evidence that dynamic and static stimuli are treated as two distinct categories. They presented participants with a visual search task, preceded by an irrelevant cue. This cue only affected subsequent search if both cue and search target were either static or dynamic. No cross-over between static and dynamic was found, suggesting that attention reflects the brain-physiology, in the sense that it is organized along a broad division between static and dynamic. These findings suggest that for the visual system a dynamic object is easy to discern when it is surrounded by static objects, since dynamic and static items are so different (physiologically speaking). Interestingly, this would also predict the reverse, since dynamic and static objects are so different, a static object should also be easy to discern when it is surrounded by dynamic objects.

However there is also evidence that not all dynamic stimuli get processed in essentially the same manner. Obviously, the final percept of motion versus flicker is different. This suggests that somewhere later in the process the brain has to distinguish between these inputs. Indeed, after V1 the processing of blinking and moving stimuli seems to diverge. The processing of moving stimuli proceeds to V5/MT, whereas the path of onset-processing is less clear. Some studies suggest that onsets are also processed in specialized areas of MT (O'Keefe & Movshon, 1998), whereas other research indicates that the lateral intraparietal sulcus is essential for processing abruptly appearing objects (Gottlieb, Kusunoki, & Goldberg, 1998). Furthermore, patient-studies have shown that some lesions cause impairment to motion-perception, but not flicker-perception and vice versa (see Vaina, Makris, Kennedy & Cowey, 1998; Vaina, Cowey & Kennedy, 1999; Zihl, von Cramon & Mai, 1983). Importantly, in chapter two of the thesis, we report evidence that for detecting a static object in a dynamic environment, the visual system does *not* treat all dynamic objects alike (the data suggests a split between onsets and moving stimuli), suggesting that the visual system may not just employ a broad division between static and dynamic.

Static among dynamic

Altogether, the fate of a dynamic object against a static background is straightforward, but the reverse - a static target among dynamic distractors– is less clear. Can we easily detect a static item because it is so different from dynamic objects? Or are we predisposed in such a way that we cannot avoid attending to dynamic stimuli? The latter suggestion seems more in line with the absolute bottom up model, whereas the relative bottom up model would state that in a fully dynamic environment static items should actually become more attractive (since they are now the unique ones). The study of McLeod et al. (1988) supports the relative bottom up model. They showed that a non-moving item among moving items can be found independent of the number of moving distractors, although search was still somewhat less efficient than for a moving target among stationary distractors. This appears to suggest that a static item may indeed be efficiently found among dynamic items. However, it remains a question as to whether motion represents a special case or whether in general static items can be efficiently discriminated from dynamic items. Pashler’s (2001) research may shed some light on this issue. He presented participants with a visual search task in which they searched through one subset, while another to-be-ignored subset of items was blinking on and off (i.e. participants searched through all the red items, while the green items were blinking). Contrary to Pashler’s expectations, the distracting subset was more easily ignored when it was blinking than when it was static. Thus, his study indicates that blinking distractors might be ignored easily, allowing efficient search of a static target. However, Pashler’s study has some caveats. First, the to-be-searched subset had a unique color. Therefore, any effect might be due to the unique color rather than to the blinking distractors. Second, Pashler did not manipulate set size. Therefore, it is not clear whether blinking of the distractors speeded search (and thus attentional deployment) or factors unrelated to search (such as a general alerting due to the dynamics in the display). Third, Pashler only found a reliable improvement in search when all the blinking items blinked on and off simultaneously. When Pashler made the items blink at random rates he no longer found any reliable effects. Perhaps then Pashler's results were due to common-onset grouping (where items that appear together are clustered, see Jiang, Chun, & Marks, 2002) rather than blinking.

The issue as to whether static objects can attract attention was further investigated by Theeuwes (2004), who presented participants with a static target surrounded by distractors that abruptly changed orientation (i.e. a vertical or horizontal bar among tilted bars). Theeuwes (2004) found that the static target was efficiently detected and concluded that static objects can efficiently be found among dynamic distractors in general. However, Davis and Leow (2005) argued that in Theeuwes' displays apparent motion could still play a role, since a bar that changes orientation can also be viewed as a rotating bar. Therefore, Davis and Leow (2005) suggested Theeuwes (2004) essentially found the same result as McLeod et al., namely that a static target can be found among moving distractors. Theeuwes' (2004) results still did not show that a static item can be found among dynamic distractors in general.

Thus, the issue was still unresolved. Can a static target be found in any dynamic environment, because it is unique against this background? Or is it hard to find a static target even under these conditions, because we are genetically inclined to look for dynamic objects? In this thesis, I present three studies that investigate this issue in depth. These studies show that, in accordance with the relative bottom up model, a static target can be efficiently found in various dynamic environments. Furthermore, this efficient detection is not based on an advanced motion-filter, nor on top-down settings. In Chapter two, Pinto, Olivers and Theeuwes (2006) investigate whether a static target can be found when it is surrounded by dynamic, non-moving distractors. Pinto et al. presented participants with a static target surrounded by blinking items. This set-up ensured that no form of motion could play a role. In follow-up experiments Pinto et al. excluded possible confounds from luminance, long-range apparent motion and temporal grouping. With all possible confounds eliminated, the static target was still efficiently detected, implying that a static target can be immediately detected among several types of dynamical distractors. In Chapter three, Pinto, Olivers and Theeuwes (in press-a) investigated whether, for the attentional system, searching for a static target among blinking distractors is essentially the same as searching for a static target among moving distractors. Pinto et al. found evidence that this was not the case. They showed that certain manipulations, such as making the objects equiluminant with the background, affected the search for a static object among moving but not among blinking distractors. Thus, the rejection of blinking and moving distractors cannot be entirely based on the same mechanism (otherwise a manipulation would either affect both or none of these capabilities). The results of Pinto, Olivers and Theeuwes (in press-a) again suggest that the efficient detection of a static target in a dynamic environment stems from a general rule and not a particular exception.

These results put severe constraints on the absolute bottom up model. Apparently, dynamic objects loose their attractiveness (in favour of the static object) given the right surroundings. These findings clearly support the main tenet of the relative bottom up model, namely that the saliency of any object (and thus how much it attracts attention) depends on which objects surround it.

A further question would be if attention goes to the static object only because it is salient, or also because people *want* to attend to it. In Chapter four, Pinto, Olivers and Theeuwes (in press-b) investigate this issue, by employing the irrelevant-feature paradigm. In this set-up all objects except one were dynamic, and the static object was mostly one of the distractors. Therefore, participants no longer had the incentive to look for a static object (this set-up is similar to the one in earlier mentioned studies of Yantis and Jonides). Pinto et al. found that when the target happened to be static, search was more efficient than when it was dynamic. Importantly, this indicates that not only a static object can guide attention, it involuntarily captures attention when surrounded by dynamic objects.

Synchronicity: dynamics in tune among dynamics out of tune

Attention may not only be needed to distinguish objects from each other, but also to distinguish objects from the background. Previous research has revealed that *grouping principles* are essential for separating items from the background. Therefore, if attention plays a crucial role in foreground/background dissociation, it is expected that attention is needed for grouping.

Over centuries, several grouping principles have been discovered. An important non-dynamic grouping principle is grouping on symmetry, where items that are symmetric are seen as belonging together. An example of dynamic grouping is grouping by common fate, items that move together are perceived as one object (Wertheimer, 1923). Another dynamic grouping example is grouping by common onset, where items that appear together are seen as one object or surface (Jiang et al., 2002). Similarly, Sekuler and Bennett (2001) discovered grouping by common luminance change, items that change luminance at the same time and in the same direction are clustered.

A crucial question is whether attention is involved in these grouping mechanisms. The evidence so far is mixed. For instance, in visual search Olivers and van der Helm (1998) showed that focused attention is needed for detecting symmetry. On the other hand, common fate grouping and common onset grouping appear to occur without the need for focused attention, even when the to-be-grouped items are interspersed with irrelevant distractors (McLeod, Driver & Crisp, 1988). Thus, it could be argued that dynamic grouping against static backgrounds does not require attention. This is in line with both the relative and the absolute bottom up account that both state that dynamic items in static backgrounds are good in guiding attention, and therefore should be separated from their backgrounds without the need for attention.

A newly discovered grouping principle, *temporal grouping*, provides a new testing ground for attention and grouping. This is because in temporal grouping displays, both the figure *and* the background are dynamic. Figure and background typically consist of continuously and rapidly changing elements. The striking finding is that, to obtain grouping, the elements do not need to change in the same direction, nor do they need to undergo the same type of change. As long as they change at the same moment, they are grouped together.

Evidence for the existence of temporal grouping started with Fahle (1993). He found that when a group of dots changed luminance out of phase with the surrounding dots, participants saw these dots as one group. He concluded that the common moment of change caused the grouping. However, his results might also be explained by common onset grouping, since the target dots all appeared at the same moment, at a time when all the background dots were switched off.

Lee and Blake (1999) continued to investigate grouping solely based on temporal information, by presenting participants with a field of Gabor patches, each of which contained a randomly oriented grating moving in a direction perpendicular to this orientation. At random moments in time, each Gabor patch could change motion direction (by 180°). In the background these motion flips were uncorrelated, but for a central rectangular region the patches changed motion direction in synchrony. Thus, in the target rectangle the motion *direction* of the gabor patches was uncorrelated, only the *moment* that the motion direction changed was synchronized. Participants had to determine the overall orientation of the target rectangle (horizontal or vertical) and could do so almost without errors. Lee and Blake (1999) concluded that temporal information alone is sufficient to segment spatial regions from their background (see also Aslin, Blake, & Chun, 2002; Guttman, Gilroy, & Blake, 2005).

Thus, temporal grouping appears to be a grouping principle that the visual system employs. However, it is unclear if this principle requires attention. Reasoning along the lines of the absolute bottom up account would suggest that it does. None of the elements by itself has any unique properties, thus when it comes to absolute qualities none of the items could serve as a base for attention. Since groups or objects that are segregated pre-attentively from their surroundings can serve as a base for attention, this would imply that the temporal group cannot be segregated pre-attentively. Conversely, the relative bottom up model predicts the opposite. According to this model attentional distribution is based on temporal and spatial comparison. Temporal comparisons underlie temporal grouping, and thus temporal groups could serve as an attentional base (and thus be construed pre-attentively).

In Chapter 5, Pinto, Olivers and Theeuwes (in press - c) investigated this issue by employing a visual search task. They presented participants with a visual search task that encompassed several temporal groups. If these temporal groups require attention to be perceived, then it is expected that the more temporal groups there are, the longer it takes participants to respond. Pinto et al. found that the number of temporal groups did not affect reaction times. This suggests that temporal grouping occurs without attention. Thus, Pinto et al.'s findings provide another context - figure-ground segregation - in which the relative bottom up model seems to fit the data better than the absolute bottom up model.

The two dynamic situations outlined so far mostly compared the relative bottom up model to the absolute bottom up model. In these dynamical settings, the relative bottom up model fared better. However, in these situations the bottom-up models have hardly been compared to the top-down model. In the next section we will pitch the bottom up models against the top-down model by investigating what the influence of display changes between trials is on attentional distribution.

Dynamic changes between trials

Changing search displays

An item which carries a unique feature on one dimension (e.g. an object with a unique color) is referred to as a singleton (Pashler, 1988). According to the bottom-up view singletons capture attention, irrespective of top-down settings (e.g. Theeuwes, 1991, 1992). If this claim is correct, then there is at least one type of stimulus, namely a singleton, that defies top-down control over attention. This then would, at least partly, falsify the top-down view. Therefore, Yantis and Egeth (1999) have investigated whether there are top-down factors that can modulate attentional capture by singletons. In their study participants performed a difficult search task, for instance the search for a vertical bar among tilted bars. However, sometimes the target carried a unique color, which turned the target into a singleton (the unique color could also be attached to a distractor).

Yantis and Egeth manipulated how often the target and singleton coincided, making it more or less beneficial for the subject to attend to singletons. They found that when the target was rarely a singleton (e.g. on 20% of the trials) search was inefficient even on the trials where the target happened to be a singleton. However, when the target was often a singleton (e.g. on 80% of the trials) search was aided when the target carried a unique color. Their findings suggest that participants only used color information when it was beneficial (and therefore supposedly wanted to use this information). This would mean that attentional capture by singletons is under top-down control.

Another possible example of top-down influence on attentional capture by singletons comes from the studies of Theeuwes (1991, 1992). In these studies participants searched for a shape singleton (e.g. a diamond among circles) and sometimes a color singleton would be present as well (e.g. one of the circles would be red, while all the other objects would be green). In Theeuwes' (1991) experiments participants always searched for a shape singleton, but they did not know beforehand which shape singleton. That is, they searched for the unique shape, either a diamond among circles or vice versa. In this case Theeuwes found that participants were, on average, 150 ms slower when a color distractor was present. In contrast, in Theeuwes' (1992) study, participants did know the exact shape they were looking for. That is, within one block of trials, they would either always search for a diamond among circles or vice versa. In this study a distractor cost of only about 20 ms was found. Thus, these two studies combined seem to corroborate Yantis and Egeth's findings. In this case, it seems that knowing the exact shape of the target beforehand reduces attentional capture.

Importantly, an alternative explanation of the Yantis and Egeth, and Theeuwes studies is possible. This explanation would rely on dynamic changes between displays, allowing for *intertrial priming*, rather than static knowledge about the singleton which remains constant across a block. The importance of intertrial priming in search tasks has been demonstrated by Maljkovic and Nakayama (1994). They found that when a target carries a feature that it also carried on a previous trial, it is found faster. Thus, if a target is red on the current trial, and it was also red on the previous trial, reaction times are faster, than when the target was green on the previous trial. Such priming effects are not limited to one trial back. Reaction times were even affected by what the target color was up to 7 trials before the current trial (i.e. participants were faster if the current target color matched the target color seven trials before the current trial). Priming seems not to be under top-down control. First, participants have no recollection of the target color more than one trial back. This makes it unlikely that they actively searched for the color previously associated with the target. Second, when participants are induced to search for a specific color, priming effects are not modulated. For example, if participants know that on most trials the target alternates between colors (e.g. from green to red), they should search for another color than the target just had. However, even in this case, participants are faster when the target color repeats than when it changes. Thus, the priming of the target color seems not to be affected by what participants expect or look for (Maljkovic & Nakayama, 1994). Similar priming effects are found for distractors. Features associated with the distractor are automatically inhibited on subsequent trials. Thus if a distractor was red on a previous trial, then participants will be less inclined to attend to red on the current trial. Note that when a color that was previously associated with the distractor now becomes associated with the target (or vice versa) priming is at its strongest. In that situation participants are maximally slowed down due to target-distractor priming.

Perhaps then, Yantis and Egeth's (1990) results were not due to overall knowledge, but to priming on the trial level as argued by Olivers and Humphreys (2003). Increasing the coincidence, increases the number of times the target has the same color as on the previous trial, thus increasing target-target priming. Furthermore, it reduces the number of times that the target on the current trial has the color of a distractor on the previous trial, reducing target-distractor priming. Thus, by increasing the coincidence between target and singleton, search may be made more efficient due to priming, rather than to knowledge.

Similarly, the difference in distractor costs between Theeuwes (1991) and Theeuwes (1992) studies might also be due to changed intertrial priming relations rather than distractor knowledge. In a pure block (i.e. when the shape of the target would remain the same throughout a block) the shape of the distractor on the current trial would always be the same as the shape of the distractor on the previous trial, maximally suppressing the distractor through distractor-distractor priming. However, in a mixed block (i.e. when the shape of the target and of the distractors could vary from trial to trial) the shape of distractor on the current trial was not always the same as the shape of the distractor on the previous trial. Furthermore, sometimes the shape of the distractor on the current trial was the same as the shape of the target on the previous trial, maximally enhancing the attractiveness of the distractor through target-distractor priming. Thus, perhaps priming, rather than knowledge, caused more suppression of the distractor in a pure block than in a mixed block.

In Chapter six, Pinto, Olivers and Theeuwes (2005) investigated this issue. First they showed that the distractor effect in the mixed block is indeed larger than the distractor effect in the pure block. Second, they compared repetition trials (i.e. when the shape of the target was the same on the current and the previous trial) in a mixed block to trials in a pure block. Note that in a pure block, all trials are repetition trials. Thus, from a priming perspective repetition trials in the mixed block are largely the same as trials in the pure block (since in both cases target and distractor shapes repeat). However, from a knowledge perspective, repetition trials in the mixed block are entirely different from trials in the pure block. Since trials are randomly presented, in the mixed block participants have no way of knowing that the target will have the same shape on the current trial as on the previous trial (when on the previous trial the target was a diamond this does not increase the chance that on the current trial the target will again be a diamond). Conversely, in the pure block, participants know beforehand which shape the target will carry, since in a pure block the target shape is the same on every trial. Thus, if the reduced distractor effect in the pure block is caused by participants knowing the exact shape of the target, then there should be a larger distractor effect on repetition trials in the mixed block than on trials in the pure block. However, if intertrial priming underlies the reduced distractor effect in the pure block, then the distractor effect on repetition trials in the mixed block should be (roughly) equal to the distractor effect in the pure block. Importantly, Pinto et al. found the latter. There was no difference between the distractor effect on repetition trials in the mixed block and the distractor effect in the pure block. Pinto et al.'s findings suggest that priming and not knowledge causes the reduced distractor effect.

Summarizing, although studies have been put forward that seem to show that target or distractor knowledge influences attentional distribution, closer examination reveals that these claims may have been incorrect. Not target knowledge, but dynamic changes between displays caused these effects. This strengthens the bottom-up models. It seems that target knowledge cannot influence attentional capture by singletons, suggesting that this capture is indeed outside of the control of the subject.

General Discussion

In the current thesis we have investigated three theories on attentional allocation. The relative bottom-up model states that attention is allocated based on comparisons between stimuli. The absolute bottom-up model suggests that attention is attracted to certain types of stimuli (especially dynamic ones). The top-down model argues that not stimuli, but search goals determine where attention goes.

We have reviewed two situations. The first situation involved dynamic displays, in which the target had no property by itself that could attract attention. First, we considered a static target in a dynamic environment. Second we investigated a dynamic target that was only different from its dynamic background, by integrating stimulus information over time. In both cases the target could be efficiently found, indicating that attention could be directed to the target immediately. Both findings limit the absolute bottom-up model and show that, at least in some cases, attention is driven by stimulus properties compared (over space or time) to other stimulus properties. In the second situation we reviewed how display changes across trials influences attentional allocation towards a singleton. We found that knowledge does not affect how much a singleton attracts attention, but priming (resulting from dynamic display changes) does. All findings discussed here support the relative bottom-up model.

Bottom-up versus top-down

It should be mentioned that the bottom-up models do not claim that attention is *always* bottom-up driven. First, this would hold the intuitively absurd conclusion that people can never determine themselves what they attend to, but are slaves to whatever stimuli they are confronted with. This would make day to day operating virtually impossible. Furthermore, such a notion would not be able to explain the large body of evidence that many features can guide attention. If a subject performs a search task, and knows that the target has a certain distinguishable feature, this aids his search process (Wolfe, 1994). If attention would *always* be bottom-up driven, then attention would move towards objects, irrespective of whatever knowledge or search goals the subject has. Consequently, in such a scenario knowing what to look for should not help the search process. However, the claim of the bottom-up models is more modest. They claim that *initial* attention is bottom-up driven. Thus, the controversy of top-down versus bottom-up comes down to questions regarding initial attentional deployment. If you are confronted with a new visual scene, can you immediately control where your attention goes? According to the top-down theories you can, whereas bottom-up theories claim that this is not possible. The findings we have discussed so far support the bottom-up view. Further support for the bottom-up view comes from several studies. Van Zoest, Donk and Theeuwes (2004) presented participants with a visual search task in which two singletons were present (and various non-singleton objects), one was a distractor, the other one the target. They found that fast eye movements (i.e. eye movements that started shortly after image presentation) always went to the most salient singleton, irrespective of what the target was. Conversely, slow eye movements always went to the target, irrespective of what the most salient item was. Since eye movements are preceded by attention (Deubel & Schneider, 1996; Irwin, 1992), this suggests that attention first goes to the salient object, and then moves on to the object that matches the search goal.

The hypothesis that initial attention is not under voluntary control could perhaps receive support from the *recurrent processing* theory of awareness (Lamme & Roelfsema, 2000). According to this theory participants become aware of visual information after it has been recurrently processed. Thus, according to this theory during the first feedforward sweep (which lasts about 100ms) participants do not yet consciously see the stimulus (see Figure 1.5). Only after feedback from other cortical areas arises, participants consciously perceive the image. Intuitively it makes sense that as long as you do not see an image, you cannot voluntarily steer your attention around. Thus, it could be that the initial sweep of attention is bottom-up driven, due to limitations in visual awareness.

Since it is only the first sweep that is bottom-up driven, the bottom-up models are more difficult to falsify than just providing evidence that certain stimuli do not always capture attention. The timing becomes crucial. For instance, in the study of Folk, Remington and Johnston (1992), where they showed that a cue that does not match the search goals does not affect attention 150 ms later, the 150 ms difference between cue and target might be crucial. It could be that attention initially did go to the incongruent cue (i.e. the cue that did not match the search goal), but was already disengaged 150 ms later. This explanation is supported by a study of Theeuwes, Atchley and Kramer (2000), who found that when the the time between cue and search display was shortened RT effects of the incongruent cue could be measured. That is, when time between search display and incongruent cue was less than 100 ms, participants were slowed if the incongruent cue was at a different location then where the target would appear, and they were faster when the cue appeared at the same location as the target, indicating that attention did go to the cue.

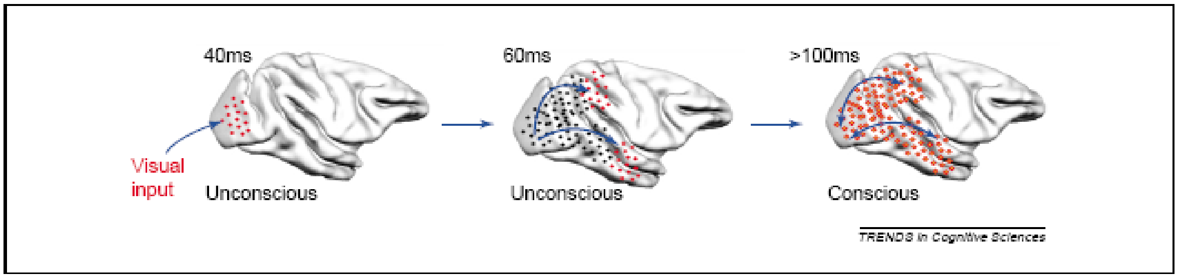

Figure 1.5. Illustration of the time course of feed-forward and feedback neural activation (taken from Lamme, 2003).

Relative versus absolute bottom-up

Although in the current thesis we present findings supporting the relative bottom-up model over the absolute bottom-up model, it should be pointed out that there is also evidence for the reverse. First, there is a search asymmetry between static and dynamic objects. A dynamic object in a static environment is found effortlessly, with search slopes indistinguishable from 0 ms/item (e.g. Mcleod et al., 1988; Watson & Humphreys, 1995). However, a static target in a dynamic environment produces a small, but significant, search slope (Pinto et al., 2006). Thus, it seems that differences between target and distractors do not entirely determine attentional deployment. After all, a static target in a dynamic environment is as different from its surroundings as a dynamic target in a static environment, but still the latter is found more easily. This search asymmetry could be due to an attentional bias. Müller and von Mühlenen (1999) found that when looking for motion-form conjunctions, participants were initially faster when they searched for a moving target than a static target. They hypothesized this was due to an attentional bias to look for dynamic items, since in everyday life attending to dynamic objects is more useful then attending to static objects (since most objects are static). Their suggestion was supported by the finding that after intensive training participants were no longer faster at detecting the motion target, supposedly since an intensive training cancelled the attentional bias. Note that even if the attentional bias can explain the search asymmetry, then this still implies that initial attention is driven by more than stimuli differences alone, and that for some reason certain stimuli are more easily attended than others.

Conclusion

All in all, a single theory on which factors determine attentional allocation seems not to yield a simple answer. Not one of the three models on attention seems to be entirely correct. Rather, it seems that the correct explanation demands a combination of these models. Attention might be allocated as follows. Initially attention is bottom-up driven, but later on top-down factors determine where attention is allocated. During the initial deployment of attention both feature differences and features by themselves are important.

Perhaps this variety of factors explains why different researchers find evidence for different models. If you measure during quick attentional deployment you will find evidence for the bottom-up models, whereas slowing participants down will yield data that supports the top-down model. Furthermore, under some circumstances properties by themselves are more important than property differences and vice versa. This leads to research finding evidence for the absolute or the relative model depending on the search task employed. All in all, this is crucial for the aim of attentional research. Rather than trying to establish which model is correct, it should be determined which model is correct under which circumstances.