The parallel representation of the objects selected by attention

Lee H. de-Wit   University of Leuven, Belgium

Geoff G. Cole   Centre for Brain Science, University of Essex, UK

Robert W. Kentridge   University of Durham, UK

A. David Milner   University of Durham, UK

The allocation of visual attention is known to be influenced by objects (B. Scholl, 2001). This object sensitivity is commonly assumed to derive from a pre-attentive stage of scene segmentation that provides a parallel representation of important structural features that can play a functional role in guiding the allocation of processing resources. In many standard "object-based attention" experiments, however, no more than two objects are ever presented. Moreover, these objects are typically presented at a predictable location for up to a second before the participant is cued to allocate attention to these objects (R. Egly, J. Driver, & R. D. Rafal, 1994). One can, therefore, ask whether many standard object-based attentional paradigms really support the notion of a parallel and pre-attentional representation. Our results, however, support the commonly held assumption that numerous objects can be maintained in parallel. Indeed, in apparent contrast to other object-based paradigms, where limits of up to (a "magic number") four are often observed, this paper found that at the least twelve objects could be maintained as potential units of selection. The results, therefore, provide evidence that the object segmentation involved in this object-based attention paradigm derives from a representation of numerous potential units of attentional selection that are maintained in parallel.

Keywords: attention, perceptual organization, 3D surface and shape perception, object recognition

Citation: de-Wit, L. H., Cole, G. G., Kentridge, R. W., & Milner, A. D. (2011). The parallel representation of the objects selected by attention. *Journal of Vision, 11*(4):13, 1–10, http://www.journalofvision.org/content/11/4/13, doi:10.1167/11.4.13.

Introduction

Following Neisser (1967), the notion of a pre-attentive stage to visual scene perception was brought to prominence by Treisman and Gelade (1980), who employed evidence from visual search to argue that scene perception could be decomposed into two distinct stages. In this model, a parallel stage of processing, in which the entire scene is segmented into basic "features," is followed by a second serial stage in which features are bound into conjoined objects. While questioning the manner in which the second stage of processing in this model might be guided by the first (Wolfe, 2003), Wolfe, Oliva, Horowitz, Butcher, and Bompas (2002) provide evidence consistent with the notion of an initial parallel pre-attentive stage of segmentation. Wolfe et al. (2002) demonstrated that if the initial segmentation process is made more difficult (by adding noise to the background) there is an increase in the latency of target detection. This increasing difficulty did not, however, increase with the number of potential targets, demonstrating that targets were not segmented one by one. Rather, the segmentation of features from their background appeared to occur in one parallel sweep for all items simultaneously. This, therefore, provides evidence for an initial all or none processing stage of segmentation in which any number of candidate (or "proto") objects across the entire visual scene are parsed simultaneously and ready to feed into the second stage of recognition.

This initial "pre-attentive" stage opens up the possibility that later attentive processing will be influenced by the manner in which the visual scene is initially organized. Exactly such a phenomenon is argued to be evident in examples of "object-based attention" (see Scholl, 2001 for a review).[1] In a number of different paradigms, the allocation of processing resources has been shown to be influenced by the initial organization of visual stimulation. These object-based attention effects are thought to reflect pre-attentive processing, because they themselves influence the allocation of attention and must, therefore, have been extracted "pre"-attentively.

A classic demonstration of object-based attention was reported by Egly, Driver, and Rafal (1994). These authors presented observers with two large rectangles, one either side of a centrally located fixation point. One end of one of the rectangles was illuminated (i.e., cued) immediately before the appearance of a target. Targets could occur either at the location of the cue, within the same object as

the cue but at the opposite end, or in the other rectangle. Replicating the basic spatial cuing effect reported by Posner (1980), reaction time (RT) was shorter when the target appeared at the cued location. More critically, however, was the observation that invalid RTs were reduced when targets were presented within the same object as the cue, relative to targets presented in the non-cued object. Egly et al. argued that the ability of these basic shapes to influence the allocation of attention demonstrated their pre-attentive extraction as objects (or "proto objects"—see Footnote 1 and Driver, Davis, Russell, Turatto, & Freeman, 2001).

There are, however, reasons to question the extent to which the visual grouping that drives Egly et al.'s cuing paradigm can truly be thought of as "pre-attentive." Goldsmith and Yeari (2003), for example, found that the initial spread of visual attention was critical to obtaining a within- versus between-object cuing advantage. More specifically, they found that the within-object cuing advantage only manifested when the participants' initial spread of attention was distributed across the scene and was abolished when the participant had to focus on the center of the display. This suggests that some degree of diffuse attention across the entire scene was required in order for objects to be constructed to a level that would allow them to influence the later allocation of attention. Goldsmith and Yeari used this evidence to explain why object-based attention effects had not been observed for endogenous cues that draw one's focus to the center of a display (Macquistan, 1997). In the present context, however, their result indicates that the visual system might have to allocate some degree of attention to an object before that object can itself influence attention.

This interactive relationship between attention and grouping accords with Driver et al.'s (2001, p. 90) contention that "most (perhaps all) of the literature on 'object-based' attention is in fact concerned with how segmentation processes constrain attentional processes, *and vice versa*" (emphasis added). Indeed, Driver et al. further note that although it represents a useful heuristic, the distinction between pre-attentive and attentive stages of visual processing represents a "gross oversimplification of biological reality."

If the segmentation required to influence attention in the widely employed Egly et al. cuing paradigm is not seen as an automatic bottom-up process but is viewed in this more interactive framework, it raises the question as to whether this object-based attention effect can play a genuine role within more complex dynamic scenes. This question is all the more pertinent if one considers the relatively trivial segmentation demands of the typical Egly et al. paradigm, in which participants are presented with two outline shapes in predictable locations. Furthermore, the within-object advantage is also assumed to occur as a result of pre-attentive processes because the rectangles are task irrelevant. However, the fact that stimuli are task irrelevant does

not guarantee that they will not be attended. Indeed, the typical Egly et al. displays are so impoverished that participants have no other competitors for the allocation of their attention, such that they are highly likely to allocate their attention to the two rectangles while waiting for cues and targets. Thus, the potential role that attention might play in extracting the required object representations and the highly predictable and oversimplified nature of the object segmentation required led us to question whether the within-object advantage would manifest with less predictable, more complex displays. Furthermore, in the classical paradigm, rectangles are presented for a relatively long time prior to the cue (typically 1000 ms), a factor that is again likely to increase the extent to which the objects are attended prior to the cue.

In the present study, we sought to investigate whether the within vs. between cuing advantage would manifest when the visual system was presented with multiple rectangles in unpredictable locations. If segmentation in the Egly et al. paradigm does indeed reflect a pre-attentive process, then this mechanism should be able to operate in parallel. Indeed, if the mechanism is truly parallel in nature, then the number of objects presented in a visual scene should not influence the strength of the "object-based" effect. If, however, some degree of serial attention has to be applied to objects before they can elicit such an effect, then an increase in the number of objects should influence the strength of the effect.

The addition of multiple object pairs in varying locations not only allows us to examine whether the visual system is able to perform more demanding levels of "pre-attentive" segmentation but also affords a comparison between the potential limits of this and other "object-based" phenomena. A number of other object-based phenomena are thought to be limited to something in the order of four objects (see Cowan, 2001 and commentaries). In the Multiple Object Tracking (MOT) paradigm (Scholl & Pylyshyn, 1999), for example, participants track not just the location of four targets but four grouped objects (Scholl, Pylyshyn, & Feldman, 2001). If the visual system is able to parse objects with unpredictable scene locations, it may be the case that there is a fundamental constraint on the number of objects that can be represented that is commensurate with the limitations seen in MOT. On the other hand, MOT may reflect a very different stage of visual selection, reflecting the number of objects that can actively be tracked after a cue. This may be very different from the number of objects that can be parsed as "potential" units of selection. A comparison between the limits in each of these "object-based attention" paradigms could, therefore, provide an important insight regarding whether these object-based attention tasks tap different stages of visual information processing.

In summary, the present work seeks to use the Egly et al. paradigm to assess whether multiple "proto objects" can indeed be maintained in parallel as potential units of

attention. To this end, the Egly et al. paradigm was adapted such that participants were presented with multiple objects at randomized locations rather than just two in the center of the display. If the influence of objects upon attention really does reflect the influence of a parallel representation derived from a pre-attentive stage of segmentation, then it should not matter how many objects the visual system is presented with: the same object-based attentional advantage should emerge. If, however, the visual system is limited in how many objects it can represent as potential units of selection, or some degree of serial selective attentional processing has to be applied to these objects before they can influence the later allocation of processing resources, then we should find a limit on the number of objects that can simultaneously be presented for a classic within-vs.-between-object advantage to be observed.

Experiment 1

Methods

Participants

Thirty-one students of the University of Durham completed Experiment 1 in exchange for course credits. Ethical permission was granted from the University of Durham Ethics Committee, and written informed consent was obtained according to the Declaration of Helsinki.

Stimuli and procedure

Each trial presented 2, 4, or 6 rectangles (see Figure 1) located at random positions. These rectangles were presented in pairs (following the logic of Egly et al.) such that the distance between a cue and a target within one rectangle was equivalent to the distance across two rectangles. The location of each pair of rectangles was determined separately, so it was possible for the rectangles to overlap. The rectangles were present for 500 ms, after which a cue was presented at the end of one rectangle of a particular pair. This cue was present for 175 ms, followed by a 50-ms interval, after which a red or green target was presented. The cue appeared either at the same location as the target (valid condition, 75% of trials), at the other end of the same rectangle (invalid within, 12.5% of trials), or at the same end of the other rectangle in that pair (invalid between, 12.5% of trials). After the participant's response, there was a 500-ms interval before the next trial. Throughout the experiment, participants were instructed to maintain fixation on the cross in the center of the display. Each participant completed 256 trials. The rectangles (0.9° by 6.4° and 0.1° thick) were presented in an 18° by 18° area in the center of the screen. Cues appeared as a lightening of the end of one of the objects, followed by targets (0.7° by 0.7°); the angular distance between the center of the cue and target on invalid trials was 5.5 degrees. Throughout the experiment, a 0.3° by 0.3° fixation cross was presented, one pixel thick, in the center of the display. Stimulus presentation was controlled from C++ using the software package
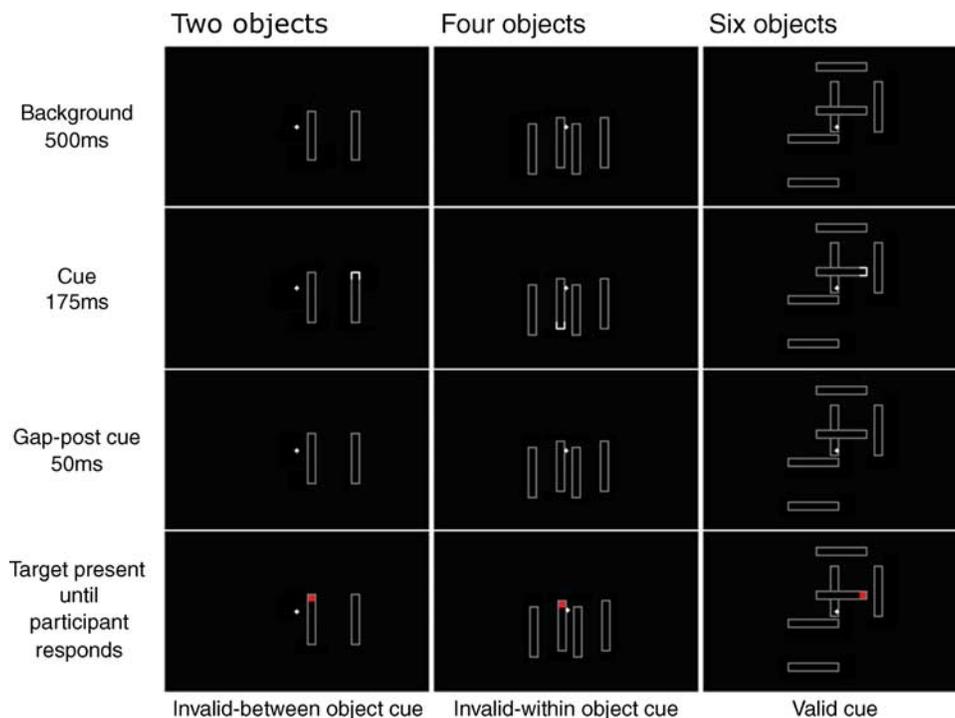

Figure 1. Stimuli illustrating the different conditions and presentation procedure for Experiment 1.

3D State. *DirectX* was used to record keyboard responses. Stimuli were presented on a Windows PC, using 17-inch monitors.

Design

We employed a 3×3 within-participants design with the number of objects (2, 4, and 6) as one factor and cue type (valid, invalid within, and invalid between) as the other.

Results and discussion

Analysis was performed on 28 of the 31 subjects tested because three of the participants performed at chance (50%) in one of the conditions. Trials with inaccurate responses or with reaction times (RTs) 2 standard deviations above or below the mean for each subject in each condition were removed from further analysis. The RTs for the valid, invalid within-, and invalid between-object trials with two, four, or six objects present are plotted in Figure 2. As expected, the RT results reveal a straightforward replication of the classic Egly et al. finding, with cued targets within the same object being detected faster than targets following cues on another object ($F(1,27) = 8.2$, $p = 0.008$). The critical question for the purpose of this experiment is whether the within- vs. between-object difference interacts with the number of objects presented. There was, however, no interaction ($F < 1$). Analysis of accuracy scores revealed no within- versus between-object sensitivity ($F < 1$) and no interaction between cue type and number of objects ($F < 1$).

The absence of an interaction between the number of objects presented and the strength of the object-based effect suggests that the visual system cannot only parse objects in unpredictable locations but can achieve this for up to six objects. Although not supported statistically, the trend apparent in the means, however, does suggest that the strength of the object-based effect begins to reduce with the presentation of more than 4 objects. Reflecting on the design of this experiment, there could have been a number of factors that may have masked a potential interaction. First, the experiment only employed a modest increase in the number of objects. If the visual system is able to automatically parse up to 4 objects, then it would still be able to parse two thirds of the objects in the six-object condition, and this could be sufficient to drive an object-based effect in this condition. Second, the locations of all rectangles were randomized independently such that they could overlap. It follows, therefore, that in the trials with three pairs of rectangles there is a greater probability that additional figure boundaries will be present between the rectangles on which the targets are presented. Thus, even if the "target" rectangles have not themselves been parsed, any potential rectangles in between might themselves act to slow the movement of attention between objects. This additional effect of the overlapping rectangles could have masked any potential decrement in the strength of the object-based attention effect in the six-object condition.

Experiment 2 sought to address both of these issues, first, by increasing the range of objects from 2, 4, and 6 to 4, 8, and 12 and, second, by stipulating that the rectangles could never overlap. In addition to these changes, the background presentation time, during which the rectangles are presented prior to the cue, was reduced from 500 to 300 ms.
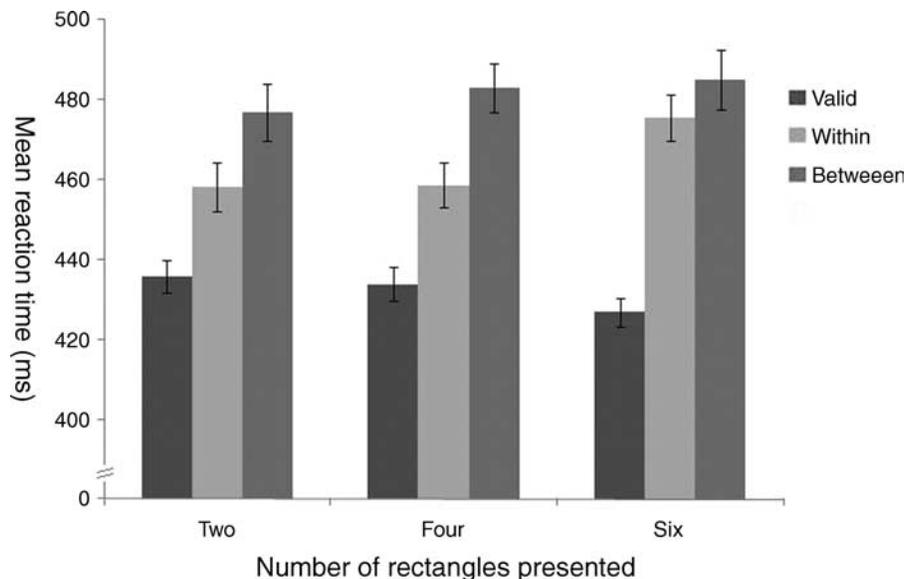

Figure 2. Reaction times (ms) and standard errors for the valid, invalid within, and invalid between conditions with two, four, or six objects present.

This reduction was employed in order to bring the presentation time of the objects prior to the cue into something more commensurate with the average fixation duration seen in natural scene perception.

Experiment 2

Methods

Except for those aspects detailed below, the methods were the same as those for Experiment 1.

Participants

Forty students from the University of Durham completed Experiment 2 either voluntarily or in exchange for course credits. Ethical permission was granted from the University of Durham Ethics Committee, and written informed consent was obtained according to the Declaration of Helsinki.

Stimuli and procedure

On each trial, the participant was first presented with 4, 8, or 12 rectangles (see Figure 3). The locations of the rectangles were distributed over a 16° by 16° area of the screen. To avoid any overlap, each pair was assigned to one of 16 locations in a virtual 4 by 4 grid across this area of the display screen (with a random offset of up to 0.35°). The rectangles were present for 300 ms, after which a cue was presented at the end of one rectangle of a given pair of rectangles. This cue was present for 75 ms, followed by a 100-ms interval, followed by a red or green target at the same location as the target (valid condition, 33.3% of
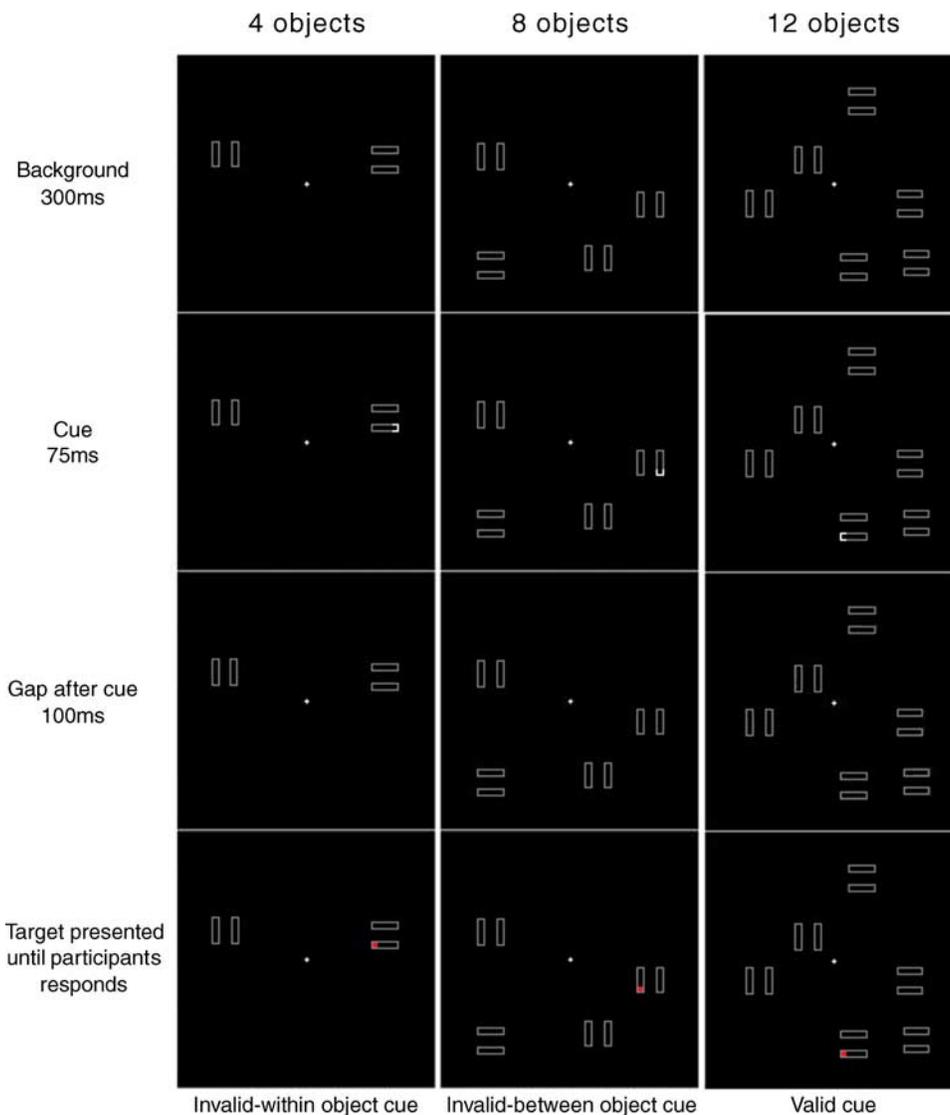

Figure 3. Stimuli illustrating the different conditions and presentation procedure for Experiment 2.

trials) at the other end of the same rectangle (invalid within, 33.3% of trials) or at the same end of the other rectangle in that pair (invalid between, 33.3% of trials). The higher ratio of invalid trials in Experiment 2 was employed to increase the number of trials contributing to the mean in each of these conditions without increasing the overall length of the experiment. After the participant's response, there was a 500-ms interval before the next trial. Throughout the experiment, participants were instructed to maintain fixation on the cross in the center of the display. Each participant completed 4 practice trials and 288 trials; the four practice trials were randomly selected from the different trial conditions. Each rectangle measured 0.7° by 2.9° of visual angle with 1-pixel-thick edges. The target square was 0.6° by 0.6° and the central fixation cross was made up of two 0.4°-long and 1-pixel-thick lines.

Stimulus presentation was controlled from C++, and *DirectX* was used to control stimulus presentation and response collection. Stimuli were presented on windows PCs, using 17-inch monitors.

Results and discussion

The reaction times for the different cuing conditions with differing numbers of objects are presented in Figure 4. Trials with inaccurate responses or with reaction times two standard deviations above or below the mean for each subject in each condition were removed from further analysis. As with Experiment 1, a repeated measures ANOVA revealed a robust difference between within- vs. between-object trials ($F(1,39) = 8.93$, $p = 0.005$). Again as in Experiment 1, this effect did not interact with the number of

objects present ($F < 1$). In addition, the accuracy of participants' responses was insensitive to cue type ($F(1,39) = 1.28$, $p = 0.264$) and again showed no interaction between cue type and the number of objects ($F < 1$).

In summary, the results from Experiment 2 show that the distinction between within- vs. between-object cuing can manifest for up to twelve objects. This suggests that twelve objects can be simultaneously represented in parallel as potential units of attentional selection. There could, however, be an alternative explanation: rather than all twelve objects being represented in parallel prior to the cue, the visual system may rapidly segment the cued object only. Thus, one could maintain that, in between the presentation of the cue and the targets, the visual system is able to rapidly allocate attention to the cued rectangles to construct the objects that then themselves influence the allocation of attention. Previous experiments have demonstrated, however, that some processing time is required before objects can be represented to the extent that they can influence attention (Chen & Cave, 2008; Law & Abrams, 2002). Those studies, however, employed somewhat different paradigms and stimuli to those used here. Thus, in order to definitively rule out the possibility that the necessary object representations could have been developed "post-cue" using our experimental procedure and stimuli, we conducted a third experiment.

Experiment 3 sought to rule out the possibility that the representations required to influence attention can be developed "post-cue" by testing whether object-based attention effects are still manifest with reduced "pre-cue" object presentation times. If the within vs. between effect reflects the selection from a set of objects maintained in parallel across the visual scene, then it should take some time for this representation to develop, and the within-vs.-between-object
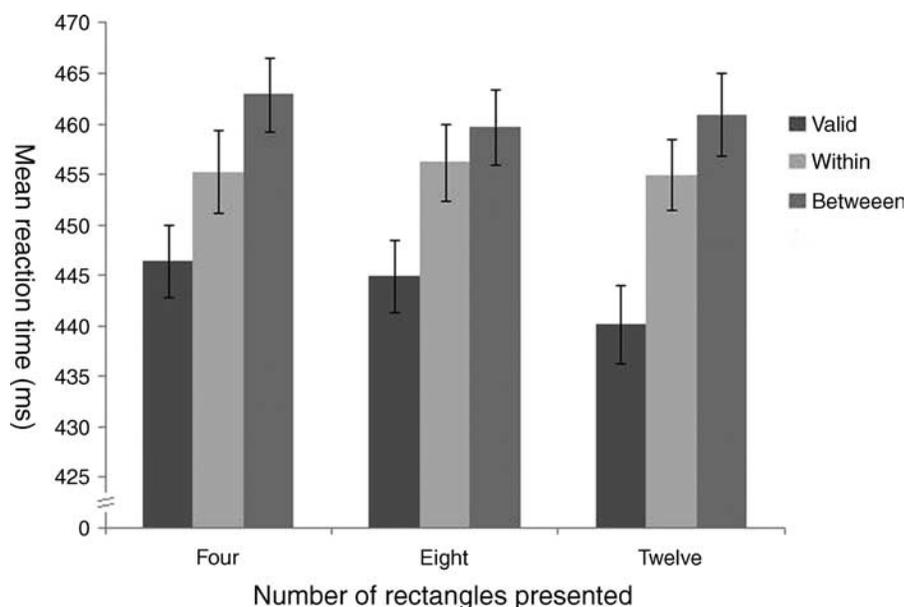

Figure 4. Reaction times and standard errors for the valid, invalid within, and invalid between cuing conditions with 4, 8, and 12 rectangles.

advantage should not be observed when the duration for which the objects are presented prior to the cue is too short.

Experiment 3

Methods

Except for the details featured below, the methods for Experiment 3 were identical to the 8-object condition of Experiment 2.

Participants

Thirty students from the University of Leuven completed Experiment 3 either voluntarily or in exchange for 4 Euros. Ethical permission was granted from the University of Leuven Ethics Committee, and written informed consent was obtained according to the Declaration of Helsinki.

Stimuli and procedure

The critical difference between Experiments 2 and 3 is that while the participant was always presented with 8 objects, these objects could now be presented for a varying length of time before the presentation of the cue. The rectangles were presented for three different durations prior to the cue: 300 ms (replicating Experiment 2), 90 ms,[2] and 0 ms (where the cue and objects were presented simultaneously). Before being presented with these rectangles, participants saw a fixation cross for 300 ms, 510 ms, or

600 ms, respectively, for the three rectangle-cue onset asynchronies (in order to keep the overall trial length equal). These new presentation durations were randomly distributed over three blocks of 216 trials (with 4 unrecorded/practice trials at the start of each block; Figure 5).

Results and discussion

A repeated measures 3 by 3 factor ANOVA revealed a main effect of presentation duration ($F(2,28) = 8.21$, $p = 0.002$) and cue type ($F(2,28) = 20.97$, $p < 0.001$). More critically, however, there was a significant interaction between cue type and presentation duration ($F(4,28) = 2.85$, $p = 0.044$). T-tests on the three within–between comparisons reveal a significant difference between within- and between-object cue–target pairings only in the 300-ms presentation condition ($t(29) = 2.18$, $p = 0.037$). There was no difference between within- and between-object invalid cues in the 0-ms condition (simultaneous presentation condition, $t < 1$) and a non-significant trend in the opposite direction in the 90-ms condition ($t(29) = 1.43$, $p = 0.165$).

The accuracy data revealed a significant effect of time ($F(2,28) = 13.3$, $p < 0.001$) such that participants respond more accurately with longer rectangle presentation times but no effect of cue type ($F < 1$) and no interaction ($F < 1$).

In sum, the results reveal a robust within-vs.-between-object advantage only when the inducing objects have been presented for 300 ms. Given that no advantage was seen at 0 or 90 ms, this suggests that between 90 and 300 ms of pre-cue stimulus processing time is required before the stimuli can be represented in a manner that allows them to act as
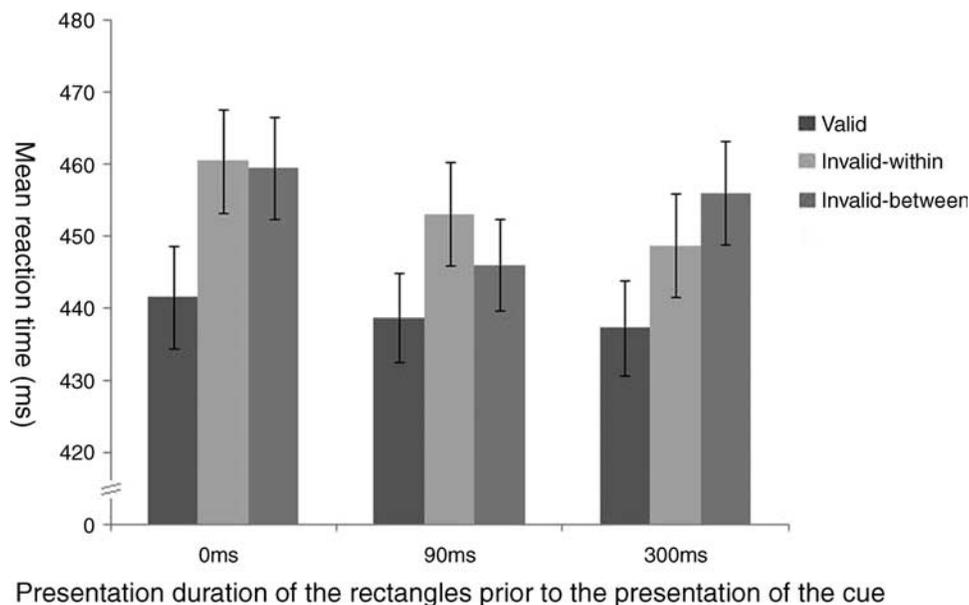

Figure 5. Reaction times and standard errors for the valid, invalid within, and invalid between cuing conditions with 8 rectangles presented either simultaneously with the cue or 90 ms or 300 ms before.

potential units of attentional selection. This effect demonstrates that the potential units of attentional selection have to be extracted prior to the presentation of a visual cue and, therefore, confirms that the multiple objects presented in this experiment have to be maintained in parallel before the presentation of the cue.

General discussion

The apparent "pre-attentive" segmentation of objects that can influence the allocation of attention has typically been demonstrated with little more than 2 or 3 objects presented at one time (Duncan, 1984; Egly et al., 1994). This fact led us to assess whether classical demonstrations of object-based attention could be applicable to scene perception where we are presented with many potential objects in unpredictable locations. Indeed, if the phenomenon of object-based attention is to play any functional role in real-world scene perception, we reasoned that it would have to select upon a representation that is maintained in parallel across an entire visual scene for multiple potential objects.

In Experiments 1 and 2, we tested whether a within-vs.-between-object advantage, found in the Egly et al. cuing paradigm, would still manifest when participants were presented with up to 12 objects. Both of these experiments provided a straightforward result, in that the within-vs.-between-object advantage did not interact with the number of objects presented. This result could be taken as evidence that (without any apparent reduction in the effect) up to twelve potential units of selection could be represented in parallel. Whether this parallel set of representations should be regarded as objects is not a straightforward question (see Driver et al., 2001, for a critical discussion regarding whether the units that influence attention can be regarded as objects). Nevertheless, our results demonstrate that whatever representations are developed for two objects can be maintained in parallel for up to 12 objects with apparently no reduction in the extent to which each unit (object) can influence attention. This, therefore, confirms a common assumption that whatever representation is developed in object-based paradigms with 2 or 3 objects can be maintained in parallel. Furthermore, it also brings the results with impoverished displays of only 2 or 3 objects a step closer to a demonstration that object-based effects like those demonstrated by Egly et al. can play a functional role in real-world scene perception where we are confronted with multiple objects.

As it stood, however, the interpretation derived from Experiments 1 and 2 could be questioned, and one could argue that, in fact, the objects did not need to be maintained in parallel, because the presentation of the cue might enable participants to very rapidly parse just a small set of objects close to the cue. Experiment 3 rules out this explanation by demonstrating that if the cue was presented simultaneously with the outline rectangles no within-vs.-between-object

cuing advantage was observed. Indeed, even if the rectangles preceded the cue by 90 ms, there was still no object-based effect. This demonstrates that objects cannot be parsed rapidly or simultaneously upon presentation of the cue. Instead, the objects have to be processed for some time (in this context, between 90 and 300 ms) prior to the cue, in order for them to be represented such that they can influence attention (consistent with prior demonstrations: Chen & Cave, 2008; Law & Abrams, 2002).

We should be clear that the focus of this article centers on the question of whether objects are maintained in parallel as potential units of selection. We cannot make any strong claims regarding the nature of the process by which these objects are extracted, a process that is in itself often assumed to occur in parallel (consistent with Davis & Driver, 1998). It is already clear that the process of extracting the objects that influence attention does not operate in a purely automatic or stimulus-dependent manner. Rather one's previous experience with a set of stimuli can strongly influence how they are organized as objects (Chen & Cave, 2006; Watson & Kramer, 1999). Indeed, as reviewed in the Introduction section, it also appears that some degree of "distributed" attention across an entire scene is required before objects are parsed to the level at which they influence attention (Goldsmith & Yeari, 2003). Both these factors could lead one to question whether the process of extracting the object representations that influence attention can operate in parallel. On the other hand, however, one could argue from the rather long durations associated with measures of attentional "dwell time" (Duncan, Ward, & Shapiro, 1994) that it is unlikely that the object extraction occurring within 300 ms for up to 12 objects in Experiment 2 could reflect a serially allocated attentional process. Given the arguments pro and con however, it would clearly be preferable to test directly whether the object representations are extracted in parallel by exploring the presentation time required to observe a within-vs.-between-object advantage for differing numbers of objects. We know that this time must be between 90 and 300 ms for 8 objects, but it could vary for differing numbers of objects. If the presentation time required for differing numbers of rectangles to generate a within-vs.-between-object advantage differs, this would provide strong evidence that they are not extracted in parallel. Such a conclusion would, however, not be incompatible with the current claim that, after these objects are extracted, their representations can be maintained in parallel.

Of course while the current experiment finds no reduction in the strength of the within-vs.-between-object advantage for up to 12 objects, one could still question whether there is an upper limit on the number of objects that can be represented in parallel. Yet even as this result stands, it provides an interesting contrast to other object-based paradigms, such as Multiple Object Tracking (MOT), which reveal a limit of something in the range of 4 or 5 objects. The manipulation of the number of objects in the current experiment and in the MOT paradigm is

clearly different however, because although the present research has highlighted that many objects can be represented simultaneously as potential units of selection, only one of those objects is selected. This is in clear contrast to MOT where up to four or five objects can be selected/tracked. Thus, while the current paper shows that multiple objects can be maintained in parallel, one could ask how many of these objects could be simultaneously cued such that targets presented on one of them would still lead to the within-vs.-between-object advantage. Indeed, while the Egly et al. paradigm and MOT are often discussed under the common umbrella term "object-based attention," a more direct comparison of the number of objects that can be selected in each paradigm could provide an important indication regarding whether or not these paradigms really reflect common underlying mechanisms. It is also pertinent to recall (as discussed in the Introduction section) that although it has already been demonstrated that something in the order of 4 objects can be tracked in parallel within the MOT paradigm, this limitation does not necessarily sit in conflict with the current result. Rather, it highlights how our report on the influence of the number of rectangles presented and the manipulation of the number of objects to be tracked may tap different stages of representation and selection. More specifically, the limitations seen in MOT pertain to a post-cue selection/tracking and, therefore, do not address the question of how many objects can be parsed and represented as potential units of attentional selection.

In summary, this research clearly highlights some important questions for future research, regarding both the potentially parallel nature of the processes involved in extracting the object representations that can influence attention and regarding how many objects can simultaneously be cued and still generate object-based effects in the Egly et al. paradigm (particularly in comparison to the limit of 4/5 objects seen in MOT). As it stands, the current research allows us to conclude that the within-vs.-between-object advantage in the Egly et al. paradigm reflects selection from a stage of representation that can simultaneously maintain multiple units of attentional selection in parallel across a visual scene. Although such a representation has been implicitly assumed, demonstrating its role in this paradigm provides an important step toward proving that the object-based attention effects apparent with simplified displays can scale up to, and potentially play a functional role in, the allocation of processing resources in real-world scene perception.

Acknowledgments

Footnotes

[1]It should be noted that the term "object" in the object-based attention literature often has a different meaning to that introduced in the first paragraph of this paper, because while objects can be regarded as a conjunction of different features, in many instances it is very likely to be the grouping along one feature dimension (rather than conjoined objects) that influences attention. In the current context this feature is shape information pertinent to perceptual segmentation and grouping (see Driver et al., 2001 for a critical discussion of the use of the term "object" in the context of the attention literature).

[2]This duration was actually 7 frames on a 75-Hz monitor, thus 90 ms is rounded down from 93.333 ms. This intermediate (90 ms) time was selected because we feared that a basic spatial attention effect might not be apparent in the simultaneous condition (because the transient onset of the cue is combined with the onset of the rectangles), and we felt 90 ms would be a sensible time at which to still expect a clear spatial attention effect but not enough time to have extracted the needed object representations.

References

Chen, Z., & Cave, K. R. (2006). Reinstating object-based attention under positional certainty: The importance of subjective parsing. *Perception & Psychophysics, 68,* 992–1003.

Chen, Z., & Cave, K. R. (2008). Object-based attention with endogenous cuing and positional certainty. *Perception & Psychophysics, 70,* 1435–1443.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Davis, G., & Driver, J. (1998). Kanizsa subjective figures can act as occluding surfaces at parallel stages of visual search. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 169–184.

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

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

Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature, 369,* 313–315.

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

Goldsmith, M., & Yeari, M. (2003). Modulation of object-based attention by spatial focus under endogenous and exogenous orienting. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 897–918.

Law, M. B., & Abrams, R. A. (2002). Object-based selection within and beyond the focus of spatial attention. *Perception & Psychophysics, 64,* 1017–1027.

Macquistan, A. D. (1997). Object-based allocation of visual attention in response to exogenous, but not endogenous, spatial precues. *Psychonomic Bulletin & Review, 4,* 512–515.

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

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

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

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

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

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

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

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences, 7,* 70–76.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research, 42,* 2985–3004.