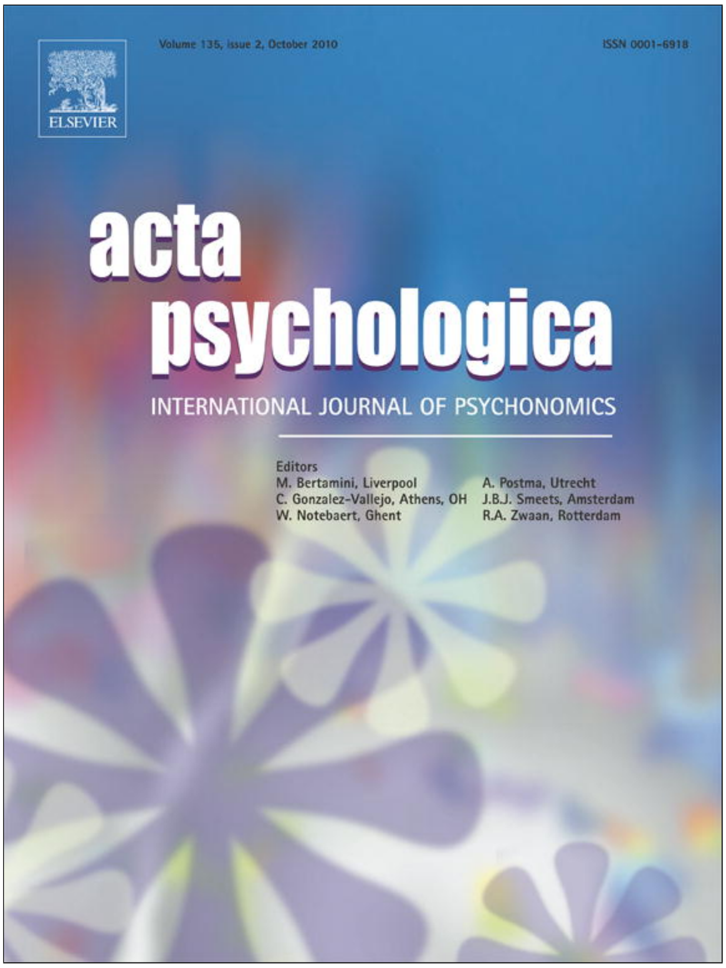
Volume 135, issue 2, October 2010
ISSN 0001-6918
ELSEVIER
acta psychologica
INTERNATIONAL JOURNAL OF PSYCHONOMICS
Editors
M. Bertamini, Liverpool
C. Gonzalez-Vallejo, Athens, OH
W. Notebaert, Ghent
A. Postma, Utrecht
J.B.J. Smeets, Amsterdam
R.A. Zwaan, Rotterdam

Acta Psychologica 135 (2010) 133–139

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Author's Reply

Top-down and bottom-up control of visual selection: Reply to commentaries

Jan Theeuwes *

Vrije Universiteit Amsterdam, The Netherlands

I would like to thank the commentators for the time and energy they put into reading and commenting to my article (Theeuwes, 2010). This type of criticism helps me to clarify misunderstandings and forces me to describe my ideas more precisely. It also brings new literature to the table and puts the previous data in a different perspective. I have organized the comments of the commentators into separate sections and addressed the issues that were commonly raised.

1. Defining top-down and bottom-up control

In my target article I defined top-down selection as an active volitional process, while bottom-up selection is thought to be passive and automatic. To sharpen these definitions, I suggested that we can only speak of top-down control when a person decides at will to select a particular feature (say all objects that are red) from the environment. In their commentary Egeth, Leonard, and Leber (2010) argued that such a strict definition of top-down control was too extreme. I think my definition was not too extreme. It is important that I added the notion "at will" to the definition of top-down control because traditionally this is the way we think about top-down control. The best example for top-down control *at will* is the classic Posner cueing task in which each trial observers can choose to select one location (or object) over another (Posner, 1980). For example, on one trial people select an object on the right side of fixation while on the next trial they select an object on the left. This is selection *at will*. Now if one needs several trials to establish an attentional set for a particular feature (say the color red) one cannot truly speak of top-down control because one needs several iterations with the environment to get the system to start selecting red "things". In fact, what happens is that processing the color red in the previous trial establishes the attentional set for the color red on the current trial. We believe that the attentional set is then established through bottom-up priming in the sense that processing red cannot be counteracted by top-down control (Belopolsky, Schreij, & Theeuwes, 2010).

The definition of bottom-up processing as a passive and automatic process received even more criticism. Kristjánsson (2010), Egeth et al. (2010), De Fockert (2010), Eimer and Kiss (2010) and Müller et al. (2010) do not necessarily subscribe to the idea that priming is bottom-up in nature. The general consensus is that bottom-up processing is completely driven by salience which by the commentator's definition is a physical characteristic of a stimulus as it exists in the outside world. I argued that prior history with a particular stimulus, as for example found in intertrial priming, may change the "appearance" of a stimulus. For example, Desimone (1996) suggested that repeated processing of a stimulus produces a "sharpening" of its cortical representation, possibly making it more salient within its environment. Bichot and Schall (2002) showed that repeating a stimulus changed responses of neurons in the frontal eye field (FEF), a region that has been implicated to be the neural substrate of the salience map (Thompson & Bichot, 2005). Our notion is that the salience is not solely defined by the physical appearance of a stimulus in the outside world, but depends on its representation in the salience map. As noted, the processing of a stimulus leads to a change in representation of that stimulus in the salience map and this change occurs independently of top-down intentions. To appreciate what this implies: if observers process an object with a red color on a given trial, on the next upcoming trial they will be biased to process red objects even when they are told to look for an object with a different color (Theeuwes, Reimann, & Mortier, 2006). Processing the color red has sharpened the cortical representation of red such that on the next occasion (i.e., the next trial) when this color is encountered again, it appears more salient. Because it is more salient, it is more likely to be selected and this occurs independently of the intentions of the observer. In this sense, bottom-up processing is considered to be automatic and passive, not sensitive to top-down set.

In light of this discussion, Eimer and Kiss (2010) specifically question whether priming which we considered to be bottom-up and automatic is in fact conceptually different from task-set contingent attentional capture as defined by Folk, Remington, and Johnston (1992). This is an interesting interpretation. However, I do not agree with Eimer and Kiss that contingent capture is necessarily top-down in origin. We (Belopolsky et al., 2010) have shown that contingent capture is not necessarily top-down because observers cannot change their top-down settings on a trial by trial basis (as they for example do in a Posner cueing task). In fact, in that study, selection was determined mainly by what was selected on the previous trial, suggesting that what looks like top-down selection is in fact simply the result of bottom-up priming. Moreover, along with Folk et al., Eimer and Kiss (2010) suggest that contingent capture is *involuntary* because "*once a top-down control setting is established (usually via task instructions at the start of the experiment), singletons that match this setting attract attention even though they are known to be task-irrelevant and thus unrelated to current selection intentions*". (p. 100). Even though in a strict sense such a

* Dept. of Cognitive Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Tel.: +31 20 444 8790; fax: +31 20 444 8832.
E-mail address: J.Theeuwes@psy.vu.nl.

0001-6918/$ – see front matter
doi:10.1016/j.actpsy.2010.07.006

definition is correct (and adhered by Folk et al.), it is somewhat odd if its mechanics are carefully analyzed. If I am looking for something red, and just before I have to start searching for red, another display flashes in front of my eyes with another object that has the same color as the one I am looking for, why is it so surprising that attention goes to the location that contains the color red? And why is this type of selection called involuntary? Clearly there is nothing involuntary about it; the observers simply select in a top-down manner the feature that is fully consistent with their task. This type of selection has nothing to do with *exogenous* attentional capture. To illustrate this point, in a recent study we showed that a matching cue (a red cue while looking for a red target) in a Folk-like paradigm does not generate inhibition of return (IOR) while the non-matching cue (an abrupt onset while looking for a red target) does generate IOR (Schreij, Theeuwes, & Olivers, 2010). Because IOR is considered to be the marker of exogenous attention (see e.g., Theeuwes & Godijn, 2002), we concluded that a matching cue in a Folk-like paradigm does not capture attention, but rather is processed in a more voluntary fashion. Only when the cue does not match the top-down set one finds IOR suggesting that only in this condition one can speak of true exogenous attentional capture.

In his commentary, Rauschenberger takes an even more radical stance. He argues that the old dichotomy of top-down and bottom-up processing should be abandoned. He suggests that the visual system is never idle (except maybe after awakening from a coma or from a sleep) and that the perception–action system of a living awake organism is so rich that one will never experience true bottom-up capture. I do not agree with this claim. In everyday life there is quite a lot of bottom-up capture even though we may not always be aware of this. As we have shown, people often move their eyes to a completely irrelevant distractor present in the visual field, often without awareness (Belopolsky, Kramer, & Theeuwes, 2008; Theeuwes, Kramer, Hahn, & Irwin, 1998). To illustrate that we do not experience attentional capture only after a coma or a deep sleep: while watching the world cup soccer on TV, our eyes often get captured by the strong luminescent billboards flashing along the side of the pitch. Our goal is to watch the ball, but our attention and eyes may get captured by this distracting information. Related to this point is Rauschenberger's argument (2010) that salience can only live in the world of phenomenology. He suggests that salience is a subjective experience of a particular stimulus quality and therefore salience cannot play a role during the initial feedforward sweep of information through the brain. Indeed, it has been claimed that for awareness reentrant processing is necessary (Lamme & Roelfsema, 2000). However, I do not see why for salience one has to assume awareness. We have repeatedly shown that salient events that people are not aware of can capture attention (Mulckhuyse & Theeuwes, in press) or the eye (Belopolsky et al., 2008). Salient events may go unnoticed even though selection is affected.

2. Priming in visual search

Several commentators indicate that the findings regarding priming provide a serious challenge to my account. There are several ways in which this is spelled out. The most prominent argument regarding priming is put forward by Kristjánsson (2010). His basic claim is that even though priming is not explicitly available to the observer it cannot be bottom-up (see also Ansorge, Horstmann, & Scharlau (2010), who suggests that priming is some sort of top-down contingent processing). To strengthen his argument, Kristjansson discusses the work by Wolfe and colleagues (Wolfe, Butcher, Lee, & Hyle, 2003) who also claimed that priming is an example of implicit top-down guidance. According to Kristjansson "differences in opinion" are at the heart of the matter. Indeed, I agree that the notion of priming as described by Kristjansson as a motivational, top-down phenomenon would be problematic. But I do not agree that it is all "a matter of opinion". In fact, there is compelling evidence that priming changes early processing which makes it unlikely to be the result of top-down processing. For example, Olivers and Hickey (2010) showed that intertrial priming results in latency shifts and amplitude differences in the P1 component of the EEG signal, a signal that is seen 80 to 130 ms following display onset. Obviously since priming affects visual processing so early, it is unlikely the result of top-down processing. Because priming has already an effect during the first feedforward sweep of processing (<150 ms) one has to conclude that priming is automatic, passive and bottom-up.

To further strengthen his argument, Kristjansson brings up research that shows that priming is affected by what he calls "motivational factors" such as reward. For example, Kristjánsson, Sigurjónsdóttir, and Driver (2010) found that the amount of monetary reward associated with a particular color strongly affected the priming of pop-out on the next trial. Because reward affects selection he concludes that it has to be motivational. Again, I do not agree. We recently conducted a study looking at effects of reward in our additional singleton paradigm (Hickey, Chelazzi, & Theeuwes, in press-a). This study shows that reward affects selection in a bottom-up way. For example, this study shows that if one gets rewarded for selecting the color red, one will keep selecting the color red even when we are told to select another color (e.g., the color green). A strategic top-down attempt to change the selection priority does not work; the brain chooses to select those features that were rewarded. Also, Hickey et al. (in press-a) showed that reward changed the P1 amplitude providing evidence that this is an automatic effect not modulated by top-down set (similar to priming discussed above). Moreover, the magnitude of visual bias created by reward was predicted by the response to reward feedback in anterior cingulate cortex, an area with strong connections to dopaminergic structures in the midbrain. We argued that through these dopaminergic structures reward changed the visual salience of features that were rewarded independent of strategic control (Hickey et al., in press-a, see also Hickey, Chelazzi, & Theeuwes, in press-b).

There were also several commentaries regarding one of our studies on intertrial priming (Pinto, Olivers, & Theeuwes, 2005). In that study we showed that in an additional singleton task, capture effects were much larger when target and distractor singleton changed from trial to trial than when they remained fixed over a block of trial. Even though most commentators (Egeth et al., Lamy et al., Kristjansson) attributed this effect to top-down expectancy and the engagement of feature search, we demonstrated that this increased distractor singleton effect was entirely due to intertrial priming, since the increased costs occurred only on trials in which the target and the distractor singleton swapped identity (Experiment 1) or on trials in which the target alone changed identity while the distractor singleton remained constant (Experiment 2). The commentators referred to a study by Lamy, Carmel, Egeth, and Leber (2006) which showed the opposite results of ours, which led them to conclude that expectancy regarding the upcoming target played a crucial role. The discrepancy between Pinto et al. (2005) and Lamy et al. (2006) was recently addressed in a study by Lamy and Yashar (2008). Like Pinto et al., they first presented distractor present and absent trials in pure blocks. With this design (as in Theeuwes, 1991) they replicated Pinto et al. (2005) providing evidence for our bottom-up intertrial priming interpretation. However, when using a mixed design (in which distractor present and absent trials were presented randomly with a block of trials), intertrial priming effects did not modulate distractor interference. Lamy and Yashar (2008) interpret these latter finding as evidence for top-down guidance of attention. However, I believe they have overstated their conclusion. As I have argued in my target article the size of the interference effect in singleton search is not so much determined by the amount of capture, but much more by the speed with which attention can be disengaged from the distractor location after it has been captured. If the distractor is less frequent and in case of a mixed design only present on half of the trials (as in Lamy et al., 2006 and Lamy & Yashar, 2008), the speed of disengagement will be relatively slow because a less frequent distractor introduces uncertainty. Observers have to determine each trial whether they were captured by the distractor or by the target (this is also how we interpret Geyer, Müller, & Krummenacher, 2008). This uncertainty slows down distractor absent trials in the mixed condition (when the target switches its shape), which explains that in the mixed condition distractor interference is unaffected by target repetition. In other words, because in the mixed condition, observers were slowed in distractor absent trials, distractor inference effect (which is the difference in RT between distractor present and distractor absent trials) was not longer affected by target repetition (i.e., intertrial priming).

Finally, Ansorge et al. (2010) argues that intertrial priming cannot play a role because usually the interval between trials is at least 1000 ms. According to Ansorge this interval between trials is too long to affect feedforward processing. It is not immediately clear why the interval between the trials is critical. We assume that processing a stimulus produces a "sharpening" of its cortical representation, and when this stimulus is encountered again (even with very long intervals in between the encounters) this will affect this initial feedforward processing stream as evidenced by changes in the early P1 effects of the ERP signal.

3. Top-down control of the (speed) of disengagement

According to my account, initial capture is basically automatic, driven by bottom-up saliency. However, once attention is captured to the location of the salient object, focal attention to that location ensures that the feature characteristics of the object at that location become available (i.e., whether it is a red diamond or a green square). This information is then matched to a top-down template (e.g., "*was my top-down goal to look for a green square?*"). If the object matches the top-down goal a response is given; if not, attention is disengaged and redirected to the object that is next in line with respect to salience. I assume that the time it takes to disengage attention from a location is dependent on how similar the selected object is with the top-down template. If the distractor looks like the target (e.g., it is a green square instead of a green diamond) it will take longer to disengage attention from that location than when the object does not look like the target at all (e.g., the distractor is a red circle when looking for a green square). We claim that the decision to disengage attention from a location is under top-down control; the ease with which that is possible depends on the target distractor similarity.

Critical to our disengagement hypothesis is the notion that objects – if salient enough – capture attention; yet when the object that captures our attention is not related to the target at all and looks completely different, disengagement is rapid, so rapid that in spatial cueing paradigm such as that of Folk et al. (1992) one cannot measure the cue validity effect anymore. Similarly, if disengagement is so rapid that attention only remains briefly at a location in space, one will not observe an N2pc component in the ERP signal. If so, it is clear that one cannot take the absence of an N2pc as evidence that attention never went to that location.

Even though the idea that the speed of disengagement depends on the extent to which the object selected matches that of the target seems quite reasonable and viable, several commentators argue that the evidence for a disengagement mechanism is quite meager and not convincing. Lamy (2010) concludes that even though the conjecture of disengagement eventually may turn out to be valid, there is currently not much evidence supporting it. Folk and Remington argue that a critical evaluation of the evidence suggests that the rapid disengagement account of contingent capture is at best unsupported, and at worst, unfalsifiable. Nordfang and Bundesen ask for a further specification of the temporal characteristics of attentional disengagement.

I agree with these criticisms and I also believe that more evidence is needed. It is indeed crucial to determine the temporal characteristics of disengagement even though it should be realized that time it takes to disengage is not fixed (which was suggested in Ansorge's commentary) but depends on the similarity between the target and the distractor. In some situations disengagement of attention can be extremely slow. For example, in a study in which observers had to interpret the direction of an arrow which indicated where to move attention next, we showed that it took at least 250 ms before observers were able to disengage attention from that location (Theeuwes, Godijn, & Pratt, 2004). In this study, we tried to speed up disengagement but the minimum time remained about 250 ms. Clearly, more research is needed to determine the spatial and temporal characteristics of attentional disengagement. Currently, we are running several studies to address these issues.

However, one piece of evidence brought up by the commentators needs to be discussed in more detail. The study by Belopolsky et al. (2010), which is also discussed in detail in Folk and Remington's commentary (2010), provides the first piece of direct evidence that rapid disengagement from a distractor took place. This study used the classic spatial cueing paradigm of Folk and showed that when a cue does not match the target property (and participants had a strong top-down set reminding them what they were looking for) there was spatially selective suppression at the non-matching cue location. The idea here is that rapid disengagement can lead to selective suppression at that location. In their commentary Folk and Remington argue that *"There is no independent evidence convincingly tying such effects to exogenous shifts of attention (Note that given the timing, the effect is not a form of Inhibition of Return)" p.* 104. Even though it is true that we cannot tie this effect directly to attentional capture, it is clear that this suppression (which was quite strong) has to be traced back to some attentional process. If one does not want to tie it to some form of attentional deployment, one has to assume that there is locally suppressive inhibition without attention ever went to that location. This alternative explanation (inhibition without attention) is theoretically quite implausible and is, as far as I know, not supported by any data.

Moreover, unlike what Folk and Remington (2010) claim, it is certainly possible that this suppression is some form of IOR since it has been shown that when attention is rapidly disengaged IOR kicks in 50 ms after display onset (Danziger & Kingstone, 1999). Also, Lamy and colleagues (Lamy & Egeth, 2003; Lamy, Leber, & Egeth, 2004) using Folk et al.'s spatial cueing paradigm reported suppression of color singletons even when the color cues preceded the onset target display by only 60 ms. Since it is generally agreed that suppression of a location can only occur after attention has been directed to and removed from a location, it is in fact very likely that this suppression is the result of attentional capture followed by fast disengagement.

Several commentators (Folk and Remington, Lamy) indicate that the study by Chen and Mordkoff (2007), who employed a fixed SOA of 35 ms between the cue and the target falsifies the fast disengagement claim. Granted to accommodate these results, we have to assume that observers can disengage very rapidly. However, even though on the surface the study of Chen and Mordkoff (2007) seems to be a Folk-like paradigm with a shorter SOA, it has a few potentially important differences. Unlike the Folk's et al. paradigm, both the cue and the target were simultaneously presented for a brief period of time. This means that when the cue was valid only one singleton was present in the display, while when the cue was invalid two singletons were simultaneously present in the display, causing competition between the popping out elements. This implies that the validity effect reported by Chen and Mordkoff (2007) may be due to the fact that there were two pop-out elements when the cue was invalid and one pop-out element (the target) when the cue was valid. The notion that the processes involved in this study are possibly completely different than those involved in the classic Folk et al. study fits with the overall RTs observed in Chen and Mordkoff (2007). Even though one would expect that the overall RTs in Chen and Mordkoff (2007) would be much longer than in Folk et al. (with such a short SOA there is less time to prepare) the overall RTs in Chen and Mordkoff (2007) were about 125 ms faster than in Folk et al. (1992)! Clearly, more research is needed to determine the exact underlying processes. We are currently running experiments to clarify these discrepancies.

Furthermore, the interpretation of the data of Anderson and Folk (2010) that I discussed in my target article as evidence for disengagement was criticized by Folk and Remington and by Lamy. Because the effect of the similarity between cue and target color was mainly found in the invalid condition (and not in the valid condition), in my view this must mean that observers had trouble disengaging attention from the invalid cue location when the color of the cue looks like the color of the target. Lamy argues that this reasoning is invalid because I based this interpretation only by looking at the graphs (and not necessarily on statistics). True; the graphs show a very convincing pattern which begs for an interpretation in terms of disengagement. Folk and Remington have another objection. They argue that my reasoning has to be incorrect because disengagement should begin immediately upon cue presentation because observers do not know whether the cue is going to be correct or incorrect. Even though observers may indeed start trying to disengage attention immediately, we have no idea how long it will take before this disengagement is complete, especially when the cue looks like a target. Our argument is that the typical cueing in all Folk et al. experiments (not only the Anderson and Folk, 2010 paper) represents RT costs of removing attention away from the cued location. The validity effect (faster RT for valid cues) is not the result of capture. The cue validity effect is the result of slower RTs when the cue is invalid because only then there is the need for disengaging attention from that location. As the data of Anderson and Folk (2010) show, the speed of disengagement depends on the resemblance between the color of the cue and that of the target.

Another observation in Anderson and Folk (2010) that deserves discussion is the condition in which there is absolutely no resemblance between the color of the target and that of the cue (the so called 0% condition in which the cue is green and the target is red). As in Belopolsky et al. (2010) here we see clear evidence for suppression (about 20 ms; see Figure 2b of Anderson and Folk 2010). Anderson and Folk show that in this condition disengagement is so efficient that by the time the target is presented at the validly cued location that location is already suppressed giving raise to slower RTs in the valid condition than in the invalid condition. Even though Anderson and Folk (2010) do not interpret the data as such, the whole pattern (including the suppression of the location of the non-matching stimuli) seems to be perfectly consistent with our disengagement hypothesis.

4. Top-down control of the spread of spatial attention (or: is selection location based?)

In my target article, I argue that when people focus their attention on a location in space (so called narrow attentional window), objects outside that window do not capture attention. I suggest that observers can adjust the spread of attention before display onset in a top-down way. For example, in my 1991 study I showed that when observers direct their attention to a location is space (on the basis of an endogenous central arrow) abrupt onsets elsewhere in the visual fields no longer capture attention (Theeuwes, 1991; see also Yantis & Jonides, 1990). When observers spread their attention across the visual field, irrespective of their relevance salient objects will capture attention (Belopolsky & Theeuwes, submitted; Belopolsky, Zwaan, Theeuwes, & Kramer, 2007).

The concept of an attentional window was discussed by various commentators. Egeth et al. indicate that the evidence is still quite meager, but agree that it is an interesting idea that deserves further study. Nordfang and Bundesen echo this concern and also indicate that more evidence is needed. De Fockert indicates that selection does not necessarily have to be based on location information but can instead be object based.

I completely agree that the evidence for the attentional window is not overwhelming yet. I suggested the concept of the attentional window already in 1994 (Theeuwes, 1994) and only recently there have been a few studies that seem to provide evidence for this concept. Only one of our studies has been published so far (Belopolsky et al., 2007) and another study is under review (Belopolsky & Theeuwes, submitted). Importantly, however a recent study by Hernandez, Costa, and Humphreys (2010) provided converging evidence for the notion of an attentional window using a different paradigm. So even though more evidence is needed, recent studies have provided the first pieces of evidence for the concept of attentional window.

In their commentary, Nordfang and Bundesen (2010) suggest a partial report study which in their view would immediately dismiss the idea of location having a special status in selection and the notion of true attentional capture. They suggest a partial report task in which observers have to report the red letters while viewing displays consisting of a mixture of equally salient red and green letters. Nordfang and Bundesen argue that according to my account attention should go to the most salient item which would result in equal probability of reporting red and green letters (because they are equally salient). The results are quite different. People report only red letters, a result which is fully consist with research on conjunction search (e.g., Kaptein, Theeuwes, & van der Heijden, 1995) in which we showed that people selectively search only those elements that match the color of the target. Even though it may appear that this experiment suggested by Nordfang and Bundesen is problematic for the attentional capture account, there is a problem with partial report studies of this sort. Among other things, the largest problem is that none of the elements in the display are truly salient. Because the display consists of a mixture multiple red and green elements none of the elements pop-out. The type of selection in this task is similar to what we have seen with conjunction search: it is relatively slow and inefficient with enough time for top-down recurrent processing. Such tasks do not say much about attentional capture and the extent to which there is top-down control during the first feedforward sweep of information through the brain.

De Fockert (2010) points out that the idea that a flexible attentional window implicitly suggests a special status for location information. This is correct. I adhere to the notion that location information regarding an object is special and not just like any other feature (such as color, shape, and luminance). It is the only feature that can directly affect selection (Theeuwes & Van der Burg, 2007, 2008). Selection based on non-spatial features such as color or shape is indirect and is mediated by location information. Ultimately space is used as the means to select. De Fockert points out that there is growing evidence that attention can also be object based which would question my claim that top-down selection is always location based. I do not think this is provides a real challenge. I think object-based selection is often automatic (i.e., it is hard not to select an object) and my claim here is that selection on the basis of space by varying the attentional window before display onset can be done in a true top-down way. In this respect I do not think that object-based attention challenges my account. Nonetheless, it is an interesting idea that in some cases the attentional window can and may be object based.

5. Do we need to assume search modes?

One compelling way to explain attentional capture is to assume that observers have control over the extent to which they get captured by salient events. This notion suggests that there are two "search modes"; one of these modes is called the "singleton detection mode" which is a mode in which observers allow distraction to occur. In this mode one will observe the classic attentional capture effects (Bacon & Egeth, 1994). Under the right circumstances, however, one should be able to force observers into a "feature search mode" in which observers search for a specific shape. Even though there is no independent evidence for the existence of these modes, the typical interpretation is that each time in an experiment capture is found, researchers assume that observers have chosen a singleton detection mode. Similarly, each time there is no capture, researchers assume that observers have chosen a feature search mode. I have objected against using this concept because it is circular and cannot be falsified. Instead, I explained the absence in capture in terms of a top-down set for a narrow attentional window and the presence of attentional capture in terms of a wide attentional window (Belopolsky et al., 2007; Belopolsky & Theeuwes, submitted; Theeuwes, 1994).

If one believes in search modes one question that needs to be answered is why observers would not always choose a feature search mode because it would prevent attentional capture. Egeth et al. (2010) come up with an interesting suggestion: feature search is more effortful and that is why people choose a singleton detection mode. This fits with recent data by Kawahara (2010) who showed that the default search mode is singleton detection and even when observers claim they search for a specific feature (doing feature search) they are in fact using singleton search. Singleton search seems to be the "natural" search mode and only by effortfully slowing down one may get into the so called feature search mode. Obviously, the choice of slowing down is a top-down decision; yet, the question is whether by slowing down the processes involved are the same as those involved when people made a speeded response. It is likely that early salience effects dissipate when people deliberately slow down. We have seen the same mechanism in oculomotor control. Only when people respond quickly the eyes are captured by salient singletons; if they choose to be slow and remain fixated in the middle after the distractor is presented, people can ignore the distractor. This is not surprising: if I refuse to make eye movements, it is unlikely that my eyes will get captured. Clearly, if people remain fixated in the middle, salience signals in the periphery no longer control eye movement behavior (Van der Stigchel, Meeter, & Theeuwes, 2006), consistent with the notion of a focused attentional window. As an alternative, Egeth et al. suggest that slowing down is the way to obtain top-down control. Because people choose to be more conservative, they are less likely to select the salient singleton. Even though feasible, it should be realized that by slowing down the processes involved may change. Or stated differently, by slowing down particular (fast) processes that may take place (such as the capture of attention) may no longer be represented in the RT measure. We know from the classic work on reaction time measurements when people are less likely to respond, RT does not necessarily represent all processes that are involved when the task is speeded (Pachella, 1974). In this particular case, by slowing down, preattentive parallel processing across the visual field may be circumvented while people engage in serial effortful search. The role of feedforward processing is reduced while recurrent processing plays a larger role. Therefore, we do not agree with Egeth et al. that it is just a matter of a more conservative response criterion.

6. Dimensional weighting: bottom-up or top-down?

As Müller et al. (2010) point out, their dimensional weighting account (e.g., Müller, Reimann, & Krummenacher, 2003) is very similar to my account in the sense that dimensional weighting also assumes that selection is prioritized according to the highest activation in an overall saliency map. The only – yet crucial – difference is that I assume that the salience computations are completely bottom-up and automatic, not changeable in a top-down volitional manner, while Müller et al. claim the mere top-down expectation of a stimulus can change the saliency computation of the salience map (which is an account similar to Wolfe et al., 2003).

I assume that the neurons in early visual areas coding for a particular feature (say the color red) can only change their cortical representation after these neurons have been exposed to the actual stimulus feature (e.g., the neuron needs to "cycle" to set itself for that feature). Thus, presenting a patch of a color in the center of the display before a trial, will change the representation of that color in a subsequent search display. In other words, processing the color of an object before a trial will change the salience of all elements with that color in the subsequent search display. The crucial point is that this change of salience occurs independently of top-down control and is merely the result of the processing of that color. It is believed to be automatic beyond volitional control, and to be the basis for the phenomenon of "priming". In our 2006 study (Theeuwes et al., 2006) we provided evidence for this notion. We presented the actual target object as a central cue before a trial and showed improved search for that target object independent of whether the cue was predictive or not. In other words, the cue had its effect independent of any top-down set associated with the validity of the cue. We also showed that a verbal cue (a cue saying "search for the red diamond") had no effect whatsoever (see also Theeuwes & Van der Burg, 2007). In our view this is not surprising. A verbal instruction (e.g., attend to red) cannot directly affect the "cycling" of a neuron and therefore it cannot change cortical representation. The verbal cognitive expectation to attend to a particular non-spatial feature (e.g., attend to red) cannot affect early visual areas, and therefore cannot change the salience signal.

Müller et al. (2010) argue differently. They do believe that the salience map can change *at will*, by simply expecting a particular feature. In their commentary, they bring up several arguments. One argument is that in my studies there was a small non-significant 9 ms effect of the verbal cue (Theeuwes et al., 2006). Müller and Krummenacher (2006) also found the same 9 ms but now it turned out to be statistically reliable. Similarly, we did not find reliable effects of a verbal cue in 5 separate experiments in Theeuwes and Van der Burg (2007) but as Müller et al. point out, there appeared to be a numerical advance for valid versus invalid verbal cues even though it was not reliable in any of the experiments. Müller et al. (2010) point out that "*the absence of statistically reliable evidence is not evidence of absence*" (p. 120). Granted, prove for the absence of an effect is impossible. Yet, if effects of verbal cues exist, they are small, much smaller than automatic feature-based priming effects. Finding these tiny effects of a verbal cue in a few but certainly not in all experiments does not strike me as a convincing case.

The bottom line is that there are quite some similarities between dimensional weighting and our stimulus-driven salience account. The only difference is that we believe that the weights are changed though bottom-up (intertrial) priming and not by top-down modulation. In this respect, the dimensional weighting account makes a somewhat inconclusive prediction. Müller et al. point out that dimensional weighting predicts that top-down control is never perfect. Even with the strongest top-down set possible, on some trials the salience will win the race and the distractor gets selected (see also Ansorge for a similar reasoning). In other words, Müller et al. agree with my stimulus-driven capture account suggesting that top-down control cannot overcome capture, but suggest that it does not happen on all trials. In fact, we also do not believe it occurs on all trials. In a recent ERP study, we showed that salience only wins the race when participants respond fast. When they respond relatively slow, attention was directly deployed to the target (Hickey, van Zoest, & Theeuwes, 2010). Again, as outlined before, slow responding may lead to a dissipation of the salience signal giving room for top-down control over capture (see also, van Zoest, Donk & Theeuwes, 2004).

7. Physiological evidence

Many commentaries discuss physiological evidence which may provide difficulties for my attentional capture account. In my target article, I discussed results from ERP, fMRI, TMS and single cell recording. The single cell data of Ogawa and Komatsu (2004) that I discussed are basically undisputed. This study showed that for the first 175 ms after display onset, firing rates were the same regardless of whether the animal was searching for a shape singleton or a color singleton. Only after about some 175 ms, top-down modulation began to have an effect. Nordfang and Bundesen (p. 2) suggest that these data indicate "a wave of unselective processing" (in which attentional weights are computed) is followed by "a wave of selective processing" (when the attentional weights are applied). If one assumes that applying attentional weights represents the implementation of top-down control, this notion fits well with my idea that initial selection is determined by salience (<175 ms) equivalent to Nordfang and Bundesen's wave of unselective processing which is then followed by selective top-down processing.

The ERP evidence brought up in the commentaries is in my view more controversial. Eimer and colleagues have repeatedly shown that in Folk-like paradigms, the matching cue elicits an N2pc while the non-matching cue does not (e.g., Eimer & Kiss, 2008). In my target article I indicated that this type of evidence cannot be conclusive because if attention is captured in a purely bottom-up way and immediately disengaged, an N2pc component will not show up. The reason for this is that the N2pc does not represent that actual shift of attention as previously was believed but represents the actual processing of information at the location at which attention is directed. If there is not much to process (because the cue does not match the target template) attention is disengaged rapidly and no N2pc is found. In their commentary Eimer and Kiss are dismayed by my bias in the interpretation of what the N2pc entails. Importantly, however, I came to this conclusion on the basis of ERP data which were collected by Eimer and colleagues themselves (Kiss, Van Velzen, & Eimer, 2008). This paper quite convincingly shows that the N2pc is not related to the actual capture of attention. To defend the rationale for using the N2pc in the top-down bottom-up debate, in their commentary Eimer and Kiss (2010) give several arguments in an attempt to tone down their original conclusion and finally conclude that "*there is every reason to assume that the N2pc reflects precisely this initial selection itself, and not the subsequent attentional processing*" (p. 101). Still, this is certainly not what Kiss et al. (2008) concluded. In their summary statement at the end of their paper Kiss et al. (2008, p. 248) concluded: "*the present study has demonstrated that the N2pc triggered in response to pop-out visual search targets does not reflect processes involved in covert shifts of spatial attention, but is instead linked to spatially selective attentional mechanisms that occur after such shifts are completed*". This is exactly what I meant in my target article: The N2pc does not reflect the actual shift of attention but the processing that occurs after this shift is completed. In my reading, Kiss et al.'s conclusions mean exactly that.

Finally, in two commentaries, a recent study of Zhang and Luck (2009) was brought up. Egeth et al. and Müller et al. argue that this study (which I was not aware of) provides unequivocal evidence that feature-based attention to a non-spatial feature (in this case color) can increase early sensory processing for stimuli matching the current top-down attentional set. More specifically, the study shows that feature-based attention can influence feedforward sensory activity, as reflected by a modulation of the P1 wave of the ERP signal. Observers maintained fixation while monitoring a target-colored subset of moving dots presented peripherally at an attended spatial location (left or right hemifield). In order to detect occasional luminance decrements in the instructed color, observers attended either the red or the green dot. To determine whether feature-based attention can affect early feedforward processing independent of spatial attention, a probe array in the relevant color was presented at the unattended side. The results show that when the irrelevant dots matched the attended color, there was a significant increase in the amplitude of the P1 component. Importantly, this effect occurred within 100 ms of stimulus onset clearly within the early feedforward sweep of attention. Zhang and Luck (2009) conclude that color-based attention can operate as early as, and independently from, spatial attention. On the basis of these findings, Müller et al. conclude that the Zhang and Luck data are clearly inconsistent with my notion that top-down set for non-spatial feature attention can only bias visual attention after 150 ms. Egeth et al. conclude that the Zhang and Luck data are consistent with contingent capture showing an early top-down effect of the attentional set.

On the face of it, the study of Zhang and Luck (2009) is indeed quite problematic for my account, which claims that there is no top-down control during the early feedforward sweep. However, a closer inspection of the method section (in the supplementary material) reveals that this study might not be about feature-based top-down control modulating feedforward processing but instead it is about intertrial priming. It turns out that participants were instructed to attend to a particular color throughout one half of the experiment. In other words, participants attend one color (e.g., say the color red) throughout an entire block of trials. Because participants respond to the same feature every trial, there will be strong intertrial priming which we know from our own research will affect early feedforward processing (Theeuwes et al., 2006). Moreover, we also know that intertrial priming does indeed modulate P1 (Olivers & Hickey, 2010). Therefore, it is questionable that the evidence provided by Zhang and Luck represents early *top-down* modulation as Müller et al. and Egeth et al. suggest in their commentary. In fact, the Zhang and Luck data are completely consistent with our idea that priming may "sharpen" the cortical representation of a stimulus, making it more salient within its environment. Because of its increased saliency it has an effect on the early P1 component of the ERP signal. Note that finding a modulation of the P1 in an experiment identical to Zhang and Luck in which participants would be told each trial what to attend by means of a verbal cue ("attend red", "attend green") would provide this convincing evidence for feature-based top-down control during the early feedforward sweep.

8. Conclusion

The commentators identified several critical aspects of my stimulus-driven selection account. Some aspects revolve around definitions. Can one only speak of top-down control when observers select at will? Is bottom-up processing the same as automatic processing that cannot be altered in a volitional way? Research regarding concepts such as the attentional window and top-down control of attentional disengagement is still in its infancy and more empirical evidence is needed. Physiological data may be of great importance to determine the boundary conditions of the stimulus-driven account even though it should be realized that attentional capture is brief and swift and may therefore not always be picked up by the current physiological measures. Even though research of the last 20 years or so on attentional capture has been fruitful, stimulating and fun, still many questions are unanswered. It may not be surprising that there is so much interest in this subject. The question is fundamental to us: Are we in control of selection or is the environment telling our brain what to select?

References

Anderson, B. A., & Folk, C. L. (2010). Variations in the magnitude of attentional capture: Testing a two-process model. *Attention Perception & Psychophysics*, *72*, 342—352.
Ansorge, U., Horstmann, G., & Scharlau, I. (2010). Top-Down contingent attentional capture during feed-forward visual processing. *Acta Psychologica*, 135, 123—126 (this issue).
Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*, 485—496.
Belopolsky, A. V., Kramer, A. F., & Theeuwes, J. (2008). The role of awareness in processing of oculomotor capture: evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *20*, 2285—2297.
Belopolsky, A. V., Schreij, D., & Theeuwes, J. (2010). What is top down about contingent capture? *Attention, Perception, & Psychophysics*, *72*, 326—341.
Belopolsky, A. V., & Theeuwes, J. (submitted). No capture outside the attentional window.

Belopolsky, A. V., Zwaan, L., Theeuwes, J., & Kramer, A. F. (2007). The size of an attentional window modulates attentional capture by color singletons. *Psychonomic Bulletin & Review*, *14*, 934—938.

Bichot, N. P., & Schall, J. D. (2002). Priming in macaque frontal cortex during popout visual search: Feature-based facilitation and location-based inhibition of return. *Journal of Neuroscience*, *22*, 4675—4685.

Chen, P., & Mordkoff, J. T. (2007). Contingent capture at a very short SOA: Evidence against rapid disengagement. *Visual Cognition*, *15*, 637—646.

Danziger, S., & Kingstone, A. (1999). Unmasking the inhibition of return phenomenon. *Perception & Psychophysics*, *61*, 1024—1037.

de Fockert, J. (2010). Early top-down attentional modulation in visual processing. *Acta Psychologica*, 135, 112—113 (this issue).

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, *93*, 13494—13499.

Egeth, H. E., Leonard, C. J., & Leber, A. B. 2010. Why salience is not enough: Reflections on top-down selection in vision. *Acta Psychologica*, *135*, 130—132 (this issue).

Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*, 1423—1433.

Eimer, M., & Kiss, M. (2010). The top-down control of visual selection and how it is linked to the N2pc component. *Acta Psychologica.* 135, 100—102 (this issue).

Folk, C. L., & Remington, R. (2010). A critical evaluation of the disengagement hypothesis. *Acta Psychologica*, 135, 103—105 (this issue).

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology-Human Perception and Performance*, *18*, 1030—1044.

Geyer, T., Müller, H. J., & Krummenacher, J. (2008). Expectancies modulate attentional capture by salient color singletons. *Vision Research*, *48*, 1315—1326.

Hernandez, M., Costa, A., & Humphreys, G. W. (2010). The size of an attentional window affects working memory guidance. *Attention Perception & Psychophysicss*, *72*, 963—972.

Hickey, C. Chelazzi, L & Theeuwes, J. (in press-a). Reward changes salience in human vision via the anterior cingulate. Journal of Neuroscience.

Hickey, C. Chelazzi, L & Theeuwes, J. (in press-b). Reward has a residual impact on target selection in visual search, but not on the suppression of distractors. Visual Cognition.

Hickey, C., van Zoest, W., & Theeuwes, J. (2010). The time course of exogenous and endogenous control of covert attention. *Experimental Brain Research*, *201*, 789—796.

Olivers, C. N. L., & Hickey, C. (2010). Priming resolves perceptual ambiguity in visual search: Evidence from behaviour and electrophysiology. *Vision Research*, *50*, 1362—1371.

Kaptein, N. A., Theeuwes, J., & van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1053—1069.

Kawahara, J. (2010). Identifying a "default" visual search mode with operant conditioning. *Acta Psychologica 135,* 38–49.

Kristjánsson, A. (2010). Priming in visual search: A spanner in the works for Theeuwes's bottom-up attention sweeps. *Acta Psychologica*, 135, 114—116 (this issue).

Kristjánsson, Á., Sigurjónsdóttir, Ó., & Driver, J. (2010). Fortune and reversals of fortune in visual search: Reward contingencies for pop-out targets affect search efficiency and target repetition effects. *Attention, Perception, & Psychophysics*, *72*, 1229—1236.

Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology*, *45*, 240—249.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, *23*, 571—579.

Lamy, D. (2010). Reevaluating the disengagement hypothesis. *Acta Psychologica*, 135, 127—129 (this issue).

Lamy, D., Carmel, T., Egeth, H. E., & Leber, A. B. (2006). Effects of search mode and intertrial priming on singleton search. *Perception & Psychophysics*, *68*, 919—932.

Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology-Human Perception and Performance*, *29*, 1003—1020.

Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology-Human Perception and Performance*, *30*(6), 1019—1031.

Lamy, D., & Yashar, A. (2008). Intertrial target-feature changes do not lead to more distraction by singletons: target uncertainty does. *Vision Research*, *48*, 1274—1279.

Mulckhuyse, M., & Theeuwes, J. (2010). Unconscious attentional orienting to exogenous cues: a review of the literature. *Acta Psychologica, 134,* 299–309.

Müller, H. J., & Krummenacher, J. (2006). Locus of dimension weighting: Preattentive or postselective? *Visual Cognition*, *14*, 490—513.

Müller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus and expectancy-driven effects in dimensional weighing. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 1021—1035.

Müller, H. J., Töllner, T., Zehetleitner, M., Geyer, T., Rangelov, D., & Krummenacher, J. (2010). Dimension-based attention modulates feed-forward visual processing. *Acta Psychologica*, 135, 117—122 (this issue).

Nordfang, M., & Bundesen, C. (2010). Is initial visual selection completely stimulus-driven? *Acta Psychologica*. 135, 106—108 (this issue).

Ogawa, T., & Komatsu, H. (2004). Target selection in area V4 during a multidimensional visual search task. *Journal of Neuroscience*, *24*, 6371—6382.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance and Cognition* (pp. 41—82). New York: Halstead.

Pinto, Y., Olivers, C. N. L., & Theeuwes, J. (2005). Target uncertainty does not lead to more distraction by singletons: intertrial priming does. *Perception & Psychophysics*, *67*, 1354—1361.

Posner, M. I. (1980). Orienting of attention, The VIIth Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *32*, 3—25.

Rauschenberger, R.. (2010). Reentrant processing in attentional guidance — time to abandon old dichotomies. *Acta Psychologica*, *135*, 109—111 (this issue).

Schreij, D., Theeuwes, J., & Olivers, C. N. L. (2010). Irrelevant onsets cause inhibition of return regardless of attentional set. *Attention, Perception & Psychophysics*, *72*, 672—682.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, *49*, 83—90.

Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception*, *23*, 429—440.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*, 77—99 (this issue).

Theeuwes, J., & Godijn, R. (2002). Irrelevant singletons capture attention: evidence from inhibition of return. *Perception & Psychophysics*, *64*, 764—770.

Theeuwes, J., Godijn, R., & Pratt, J. (2004). A new estimation of the duration of attentional dwell time. *Psychonomic Bulletin & Review*, *11*, 60—64.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*, 379—385.

Theeuwes, J., Reimann, B., & Mortier, K. (2006). Visual search for featural singletons: No top-down modulation, only bottom-up priming. *Visual Cognition*, *14*, 466—489.

Theeuwes, J., & Van der Burg, E. (2007). The role of spatial and nonspatial information in visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1335—1351.

Theeuwes, J., & Van der Burg, E. (2008). The role of cueing in attentional capture. *Visual Cognition*, *16*, 232—247.

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, *147*, 251—262.

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience and Biobehavioral Reviews*, *30*, 666—679.

van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 746—759.

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology*, *29*, 483—502.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 121—134.

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*, 24—25.