

Failure of the extended contingent attentional capture account in multimodal settings

Rob H.J. Van der Lubbe^{1,2} and Jurjen Van der Helden¹

¹ Cognitive Psychology and Ergonomics, Universiteit Twente, Enschede, The Netherlands

² Dept. Of Psychonomics, Utrecht University, Utrecht, The Netherlands

Received 09.09.2006

Accepted 18.05.2006

Keywords

insight, heuristics, representational change with spatial attention, multimodal, mixed-blocked

ABSTRACT

Sudden changes in our environment like sound bursts or light flashes are thought to automatically attract our attention thereby affecting responses to subsequent targets, although an alternative view (the contingent attentional capture account) holds that stimuli only capture our attention when they match target features. In the current study, we examined whether an extended version of the latter view can explain exogenous cuing effects on speed and accuracy of performance to targets (uncued-cued) in multimodal settings, in which auditory and visual stimuli co-occur. To this end, we determined whether observed effects of visual and auditory cues, which were always intermixed, depend on top-down settings in "pure" blocks, in which only one target modality occurred, as compared to "mixed" blocks, in which targets were either visual or auditory. Results revealed that uni-

modal and crossmodal cuing effects depend on top-down settings. However, our findings were not in accordance with predictions derived from the extended contingent attentional capture account. Specifically, visual cues showed comparable effects for visual targets in pure and mixed blocks, but also a comparable effect for auditory targets in pure blocks, and most surprisingly, an opposite effect in mixed blocks. The latter result suggests that visual stimuli may distract attention from the auditory modality in case when the modality of the forthcoming target is unknown. The results additionally revealed that the Simon effect, the influence of correspondence or not between stimulus and response side, is modulated by exogenous cues in unimodal settings, but not in crossmodal settings. These findings accord with the view that attention plays an important role for the Simon effect, and additionally questions the directness of links between maps of visual and auditory space.

INTRODUCTION

Sudden sound bursts or light flashes in our environment are thought to automatically attract our attention, even when they give no clue about the locus and/or moment of subsequent task-relevant events. However, Folk, Remington and Johnston (1992) argued that stimuli only capture our attention when they are relevant, that is, when they match target features (the "contingent attentional capture account"). Applied to an everyday setting, the latter account might say

something like: "Sitting at a terrace and reading a newspaper while awaiting our friend who always wears bluish jeans jackets, our attention may become easily attracted by persons wearing blue coats, but not by persons wearing yellow pyjamas" (e.g. see Simons & Levin, 1997). Recent studies focused on attentional

Correspondence concerning this article should be addressed to Cognitive Psychology and Ergonomics, Rob H.J. van der Lubbe, Universiteit Twente, Postbus 217, 7500 AE Enschede, The Netherlands Tel: +31-53-489-3585, email: R.H.J.vanderLubbe@gw.utwente.nl

effects in slightly more natural multimodal settings, in which different stimulus modalities (e.g. visual, auditory, tactile) co-occur. In the current study, we extended the line of reasoning of the contingent attentional capture account from visual to multimodal settings, and focused on the implications of recent findings and of new data for this extended view.

To study the attentional effects of sudden onsets like sounds or flashes, numerous studies applied the classical Posner paradigm (e.g. see Posner & Cohen, 1984) in which to be detected or discriminated targets are preceded by earlier onsets with an interval between onset and target (denoted as SOA: stimulus onset asynchrony) of about 200 ms. The majority of these studies revealed that manual responses are faster when targets occur near the locus of the preceding onset (i.e. at cued locations), than when they occur at other uncued locations when the onset provides no information with regard to the likely position of a forthcoming target¹. The latter effect, denoted as a positive cuing effect, has been ascribed to attention being automatically attracted to the cued location and has therefore been denoted as exogenous (stimulus-driven) orienting. Apart from this attentional capture by sudden sounds or flashes (i.e. exogenous cues), our attention may also be voluntarily allocated to a specific location (driven by our intentions) as we expect something to occur at that location. The latter type of orienting has been named endogenous orienting to mark its internal control².

According to several recent studies, exogenous and endogenous orienting affect different levels of processing (see Berger, Henik, & Rafal, 2005; Hopfinger & West, 2006; Van der Lubbe & Postma, 2005; for indirect support see Theeuwes, Kramer, Hahn, & Irwin, 1998; Theeuwes, Kramer, Hahn, Irwin, & Zelinsky, 1999). In their experiments, Berger et al. (2005) demonstrated that exogenous orienting effects of abrupt onsets were unaffected by prior endogenous cues with a validity of 80%, except when discrimination of subsequent targets was made very hard. These findings led Berger et al. to conclude that endogenous and exogenous orienting are separate mechanisms that can lead to independent orienting effects, even under conditions in which they contradict each other. Additional support for the special status of abrupt onsets comes from studies demonstrating orienting effects without awareness of the exogenous cue evoking this effect (McCormick, 1997; see also Ansorge & Heumann, 2005), suggesting that these effects require no conscious control. In sum, abrupt onsets may be argued to be special as they automatically attract attention being largely unaffected by intentions.

An alternative proposal states that task goals (i.e. intentions) actively influence what stimuli attract our attention and what stimuli do not; as a consequence, exogenous cuing effects are dependent on top-down factors and are in fact not purely exogenous. This proposal has been denoted as the contingent attentional capture account (Folk, Remington, & Johnston, 1992; Folk & Remington, 1999; Wu & Remington, 2003). In their seminal paper, Folk et al. (1992) reported that irrelevant abrupt onsets only capture attention when targets are characterized by abrupt onsets, but not when targets are defined by a discontinuity in color. A recent study by Ansorge and Heumann (2003) also raised the question whether effects of onsets depend on task goals (in line with the contingent attentional capture account). In their first three experiments, they revealed that a match between *color* of onsets (cues) and targets was at least partly responsible for the observed orienting effect. In their fourth experiment, the influence of a match between the *location* of onsets and targets was examined. One might argue, in line with the contingent attentional capture account, that onsets in the Posner paradigm not only will attract attention as they provide likely information about potential target locations but also when they indicate potential target locations that are somewhat less likely. In most experiments employing the Posner paradigm, this condition is satisfied, as there is a chance (mostly even 50%) that a target will occur on or near the location of the onset. To test the possibility that effects of onsets depend on overlap with the target location, conditions were compared in which cues on potential target locations were on some trials accompanied by cues on positions at which never a target occurred. The idea of Ansorge and Heumann was that the latter cues should be ineffective according to the contingent attentional capture account. In line with this view no disruptive effect was found³. As a consequence, it may be argued that a single intentional mechanism is responsible for orienting effects as the special status of abrupt onsets is denied.

If one adopts the contingent attentional capture account, then what are possibly the implications for settings in which stimuli of different modalities co-occur? Can an extended version of this conception account for findings in these multimodal settings? Furthermore, what are the consequences with regard to the stability of orienting effects in more realistic and complex settings? Goal of the current study was to provide new insights on this matter by employing visual and auditory cues and targets, and varying their task relevance by employing blocks in which solely visual or auditory

targets occurred, or blocks in which these targets were intermixed. Before turning to our experiment, we will focus on the significance of two recent experiments with crossmodal settings for the extended contingent attentional capture account.

Several studies have convincingly shown that there are clear effects of the loci of auditory exogenous cues on the speed of simple detection or discrimination responses to visual targets (e.g., McDonald & Ward, 2000; Schmitt, Postma, & De Haan, 2000; Spence & Driver, 1997a; Van der Lubbe et al., 2006). In a recent study by Santangelo, Van der Lubbe, Olivetti Belardinelli, and Postma (2006) targets and cues appeared in the peripheral left or right field on crossmodal units (for an example see also Fig. 1), and effects of cue modality [auditory, visual, or bimodal (both visual and auditory)] and cue location were examined in a visual discrimination task. Responses were fastest in case of auditory cues, intermediate in case of bimodal cues, and slowest in case of visual onsets. More importantly, with an optimal SOA of 200 ms, effects of cue location on reaction time (RT) were independent from cue modality. In case of auditory onsets the cuing effect amounted to 13 ms (averaged across experiments), which was not different from the effects in case of visual (18 ms) and bimodal onsets (18 ms). In principle, cuing effects with different cue and target modalities can be explained in line with an extended version of the contingent attentional capture account, as they provide information about potential target locations. However, one might additionally argue that cuing effects should be larger with visual cues than with auditory cues, as feature overlap with targets (both location and modality) is much larger, which is not what was they observed. A possible reason might be that visual cues apart from their orienting effect also masked the subsequent target. In other words, although attention was attracted to the visually cued location the cue may at the same time have decreased visibility of the target due to lateral masking, which increases for positions far from fixation. As a consequence, these findings do not necessarily disqualify the extended contingent attentional capture account (see also footnote 3).

In another study, Van der Lubbe and Postma (2005) observed that participants, who knew in advance at what location to be discriminated visual targets would appear (indicated by a 100% valid arrow), were still affected by the locus of irrelevant visual and auditory onsets (but see Theeuwes, 1991; Yantis & Jonides, 1990). The magnitude of this unimodal and crossmodal cuing effect (≈ 30 ms) appeared to be comparable

when the location at which the target would occur was unknown. These findings seem not in line with the extended contingent attentional capture account. The target location varied from trial to trial, and knowledge about the forthcoming target location additionally varied from trial to trial. As a consequence, one might argue (see Mazza, Turatto, Rossi & Umiltà, 2006) that these variations were suboptimal for an appropriate setting for task-relevant features. Nevertheless, clear effects of arrow direction were observed, indicating that participants were able to adapt to these variations, which implies at least some problems for the (extended) contingent attentional capture account.

In the current study, both cue modality and target modality were varied, with targets and cues occurring on either the left or right side from fixation. In all cases, the modality and location of the cue varied randomly from trial to trial. In one condition, the target modality was always visual ("pure visual"), and in another condition it was always auditory ("pure auditory"), and in a third condition the target modality varied randomly from trial to trial ("mixed"). According to the extended contingent attentional capture account, attentional capture of a stimulus should be a function of the match of that stimulus with top-down control settings. Therefore, cuing effects of visual cues should depend on the type of condition, being largest in the pure visual condition, smaller in the mixed condition, and negligible in the pure auditory condition. An opposite pattern may be predicted to occur with auditory cues, being larger in the mixed condition than in the pure visual condition. The pattern in case of auditory targets might be less clear as the relevance of spatial aspects is much less for pitch discrimination, although the selection of an appropriate action may be facilitated in line with the view that attention fulfils a crucial role for linking perception to action (e.g. see Van der Heijden, 1992, 2004). To examine whether effects are possibly modulated by arousing properties of auditory stimuli, we measured the exerted force of the manual responses, as arousal effects are commonly reflected in increased force amplitudes (see Jaśkowski, Van der Lubbe, Wauschkuhn, Wascher, & Verleger, 2000; Jaśkowski & Włodarczyk, 2005).

A secondary interest⁴ of the current study was the possible influence of exogenous cues on the Simon or correspondence effect (e.g. see Simon, 1990). Specifically, left or right buttons had to be pressed after targets presented in the left or right field with their locus being irrelevant. Commonly, faster responses are found in case of correspondence than in case of non-correspondence between target stimulus and response

side (the Simon effect). A few earlier studies (Van der Lubbe, Keuss, & Stoffels, 1996; Van der Lubbe & Woestenburg, 1999; see also Ivanoff, Klein & Lupiáñez, 2002) have shown some support for a modulation (a reduction or even reversal) of the Simon effect in case of peripherally cued targets. The current study may provide additional information regarding the stability and universality of these effects.

METHODS

Participants

Fourteen participants, almost all students from the University of Utrecht, cooperated in this study. They received € 7 per hour, and participated on average for 3.2 hours. None of the participants reported a history of neuronal diseases, and all had normal or corrected to normal vision and hearing. The study was approved by the ethics committee of the faculty of social sciences, and all participants signed a written informed consent. One participant was excluded due to too many errors in the tone discrimination tasks, which left thirteen participants for the final analysis (all right-handed, mean age 22.3 years, six females and seven males).

Stimuli

Stimuli were presented on three crossmodal units (21x12 cm), consisting of a sound passing 8x8 green LED display (10x10 cm) in front of a loudspeaker (see Figure 1). The units were placed at 160 cm from the participant, 29.4° to either the left or the right, or in front of the participant. Each trial started with a fixation point (0.7x0.7°) presented on the middle unit for the whole duration of the trial. One second after presentation of the fixation point, either a visual (a line; the bottom row of the LEDs lighted up: 0.2x3.1°), or an auditory cue (a burst of white noise) was presented for 100 ms, either on the left or the right unit. Two hundred ms after cue onset either a visual or an auditory target was presented at the left or right unit for 100 ms. As visual targets, triangles (2.6x1.4°) were used pointing up- or downwards, and as auditory targets, pure tones of 950 or 1000 Hz were employed. At 1450 ms after target offset the next trial started.

Procedure and design

The experiment consisted of three different blocks, the order of which was counterbalanced across participants. These three blocks were additionally repeated, thus, totally six blocks were carried out. In all blocks the

same visual or auditory cues were presented, of which the location and modality varied randomly from trial to trial. Thus, cues had no predictive value regarding the forthcoming target location.

In the "pure auditory" blocks, only auditory targets occurred at the left or right unit, of which the pitch and location varied randomly from trial to trial. A 1000 Hz tone required a left response and a 950 Hz tone required a right response. On totally 320 trials both cues and targets occurred (cue modality x cue location x target location x target type). On another 80 trials only cues (i.e. catch trials) were presented (20 times cue modality x cue location). These trials were intended to be used to estimate evoked potentials by the cues to enable correction for these potentials (but see next section). In the "pure visual" blocks, visual instead of auditory targets were presented, but in other aspects these blocks were the same as the "pure auditory" blocks. Regarding response instructions, the triangle pointing up- or down-wards required a left or right response, respectively. In the "mixed" blocks, either visual or auditory targets were presented, each on 320 trials, and on 80 trials only cues occurred. The type and modality of the target in the latter blocks varied randomly from trial to trial. The same responses were required as in the pure visual and the pure auditory blocks. In all blocks, participants were instructed to respond as fast and accurately as possible.

Apparatus and data recording

Participants were seated in a standard arm chair in a silent and darkened chamber. Two response buttons (isometric weight elements; HBM GmbH) were attached to a table in front of the participants, to measure the exerted response force. The index fingers of the left and right hand had to rest on the buttons during the conditions. A force had to be exerted by the finger on the relevant button when a target occurred. Exerted force was monitored online. In case of too forceful or too weak responses, participants were informed that they should adapt their responding. Exerted force, the electroencephalogram (not discussed because of too low signal to noise ratio) and EOG were sampled from Ag/AgCl ring electrodes by a BrainAmp amplifier (Brain Products GmbH) at a rate of 250 Hz. EOG was measured both vertically from above and below the left eye (vEOG) and horizontally from the outer canthi of both eyes (hEOG). Electrode impedance was kept below 5 kΩ. EOG and force were digitally filtered on line (TC = 5.0 s, low-pass filter of 100 Hz) by the program Vision Recorder, installed on a Pentium III

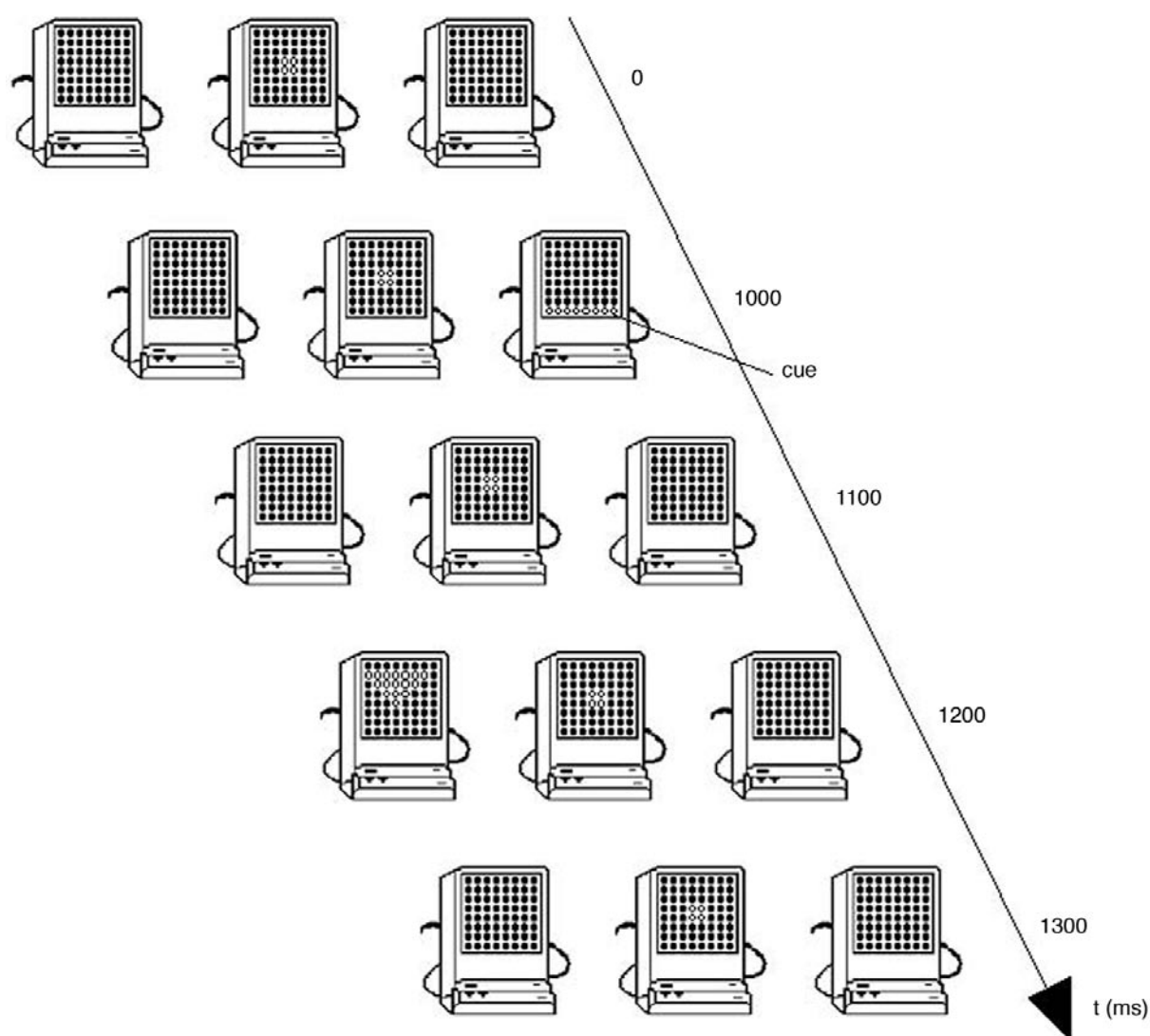


Figure 1.

An example of a unimodal trial, in which a visual cue preceded a visual target. Three units were placed in front of a participant, one to the left, one straight ahead, and one to the right. A trial started with a fixation square on the middle unit, which was followed by a cue for 100 ms on either the left or the right unit (a sound burst or a line). On most trials, a to be discriminated target (a triangle pointing up- or downwards, or a high or a low tone) was presented on either the left or the right unit at 100 ms after cue offset. The location of the cue was unpredictable with regard to the forthcoming target location. In mixed blocks cue modality and target modality varied randomly from trial to trial, whereas in pure blocks the modality of the target was fixed (either visual or auditory) but the modality of the cue varied from trial to trial.

computer. Presentation of stimuli and emission of triggers signaling the moment and type of stimulus was controlled by a CMO-module (version 3.7f, developed in cooperation with IGF, Physics department, Utrecht University, see also Van der Lubbe & Postma, 2005; Santangelo et al., 2006; Van der Lubbe, et al., 2006).

Data analysis

Trials with detectable eye movements (exceeding $60\mu\text{V}$ in the hEOG channel or $120\mu\text{V}$ in the vEOG channel) from cue onset until target offset were excluded from further analyses by using Vision Analyzer. Speed and

accuracy of responses were determined by analyses on the force channels for the left and right response buttons. Trials with forces below 0.6 N (which corresponds with a weight of approximately 61 g), or above 0.6 N on the incorrect response channel were defined as errors and were not used for the reaction time (RT) analyses. RT was defined as the moment at which the correct response channel exceeded 0.6 N . ANOVAs on RT, accuracy (percentage correct: PC), and response force were performed including the factors target modality (auditory, visual), block type (pure/mixed), cue modality (visual / auditory), cue (the location of the target relatively to the cue location; cued / uncued),

and correspondence between stimulus and response side (corresponding / noncorresponding).

RESULTS

Regarding the trials with both cues and targets, the percentages of trials without eye movements amounted to 97.4 % in case of visual targets in the pure block, 96.4 % in case of auditory targets in the pure block, and 97.0 % in case of either visual or auditory targets in the mixed block.

RT. Mean RTs and results of separate *t*-tests are presented in Table 1 and Table 2. Cuing effects, averaged across correspondence are displayed in Figure 2, and correspondence effects averaged across block are presented in Figure 3. An ANOVA with the factors target modality, block type, cue modality, cue, and correspondence revealed a main effect of block type, $F(1,12) = 79.8$, $p < .001$, with faster responses in pure than in mixed blocks (615 vs. 694 ms), a main

effect of cue modality, $F(1,12) = 12.2$, $p = .004$, with faster responses in case of auditory than in case of visual cues (643 vs. 666 ms), and a main effect of correspondence, $F(1,12) = 84.9$, $p < .001$, with faster responses for corresponding than for noncorresponding trials (638 vs. 671 ms).

The effect of block type was much larger for auditory targets than for visual targets (126 vs. 31 ms), $F(1,12) = 64.7$, $p < .001$, and this effect additionally interacted with cue modality, $F(1,12) = 10.9$, $p = 0.006$. Specifically, the delay in mixed blocks relative to pure blocks for visual targets with visual and auditory cues amounted to 13 and 48 ms (for separate tests see below), respectively, whereas this delay for auditory targets with visual cues and auditory cues amounted to 134 and 118 ms, respectively. A significant interaction between block type and cue, $F(1,12) = 4.9$, $p = .048$, and significant interactions between target modality and cue, $F(1,12) = 6.9$, $p = .022$, between block type, target modality, and

Table 1.

Mean RTs (in ms) and PCs (in %) for cued and uncued targets for each target modality, block, cue modality, and statistical results of two-sided paired t-tests.

Target modality	Cue modality	Block	Reaction times				Proportion correct			
			Cued	Uncued	<i>t</i> (12)	<i>p</i>	Cued	Uncued	<i>t</i> (12)	<i>p</i>
Visual	Visual	Pure	637	664	-2.4	.031	91.4	92.3	-0.6	.56
		Mixed	653	674	-2.2	.046	94.7	95.2	-0.6	.62
	Auditory	Pure	617	601	1.9	.087	93.0	93.6	-0.8	.46
		Mixed	651	662	-2.1	.060	95.5	95.3	0.2	.85
Auditory	Visual	Pure	600	617	-2.4	.031	89.2	89.2	0.1	.93
		Mixed	761	723	4.3	.001	88.1	88.7	-0.5	.60
	Auditory	Pure	594	595	-0.2	.88	89.8	89.8	-0.1	.11
		Mixed	715	710	-0.2	.72	91.6	90.5	1.2	.25

Table 2.

Mean RTs (in ms) and PCs (in %) for targets of which the location corresponded or not (corresponding vs. noncorresponding trials) for each combination of target modality, cue modality, and cue, and statistical results of two-sided paired t-tests.

Target modality	Cue modality	Block	Reaction times				Proportion correct			
			Corr	None	<i>t</i> (12)	<i>p</i>	Corr	None	<i>t</i> (12)	<i>p</i>
Visual	Visual	Cued	640	649	-0.8	.422	93.0	93.1	-0.1	.92
		Uncued	649	689	-3.1	.010	95.0	92.5	1.8	.09
	Auditory	Cued	618	650	-3.5	.005	95.7	92.8	2.3	.04
		Uncued	616	646	-3.7	.003	95.0	93.8	2.2	.04
Auditory	Visual	Cued	657	703	-4.5	.001	89.7	87.6	1.8	.10
		Uncued	651	689	-4.7	.001	91.0	86.9	2.5	.03
	Auditory	Cued	647	662	-2.8	.017	91.0	90.3	0.5	.65
		Uncued	683	682	-5.5	.000	90.9	89.6	0.5	.47

cue, $F(1,12) = 20.4$, $p = .001$, between block type, cue modality and cue, $F(1,12) = 13.8$, $p = .003$, and between target modality, cue modality and cue, $F(1,12) = 5.3$, $p = .04$, were found. The correspondence effect interacted with cue, $F(1,12) = 5.5$, $p = .037$, with a smaller correspondence effect for cued targets (26 ms) than for uncued targets (42 ms), and an interaction between target modality, cue modality, cue, and correspondence was found, $F(1,12) = 22.8$, $p < .001$. To facilitate understanding all these results, separate analyses were carried out for each target modality.

ANOVAs for *visual targets* revealed main effects of block type, $F(1,12) = 7.5$, $p = .018$, cue modality, with faster responses after auditory than after visual cues, $F(1,12) = 6.7$, $p = .024$ (633 vs. 657 ms), cue, $F(1,12) = 4.8$, $p = .048$ (640 vs. 650 ms), and correspondence, $F(1,12) = 13.3$, $p = .003$ (631 vs. 659 ms). A significant interaction was found between cue modality and cue, $F(1,12) = 5.6$, $p = .036$, that additionally interacted with block, $F(1,12) = 5.1$, $p = .044$, and a significant interaction was found between block and cue modality, $F(1,12) = 17.3$, $p < .001$, with a small effect of cue modality in mixed blocks as compared to pure blocks (7 vs. 42 ms). Finally, an interaction between cue modality, cue, and correspondence was observed, $F(1,12) = 6.5$, $p = .026$. No interactions were observed involving both the factors correspondence and block type, $F(1,12) < 0.1$.

Further decomposition per cue modality showed that a main cuing effect (see Fig. 2) was found with visual cues, $F(1,12) = 6.8$, $p = .023$ [independent from block type, $F(1,12) = 0.3$], without a block type effect ($F = 1.0$), but with a correspondence effect, $F(1,12) = 6.4$, $p = .027$, that additionally interacted with cue, $F(1,12) = 5.1$, $p = .043$. The correspondence effect (see Fig. 3) amounted to 9 and 41 ms for cued and uncued targets, respectively.

In case of auditory cues, a main effect of block type, $F(1,12) = 23.6$, $p < .001$, and an interaction between block type and cue was found, $F(1,12) = 7.5$, $p = .018$, the latter effect reflecting different cross-modal cuing effects in the pure and mixed blocks (see Fig. 2). Finally, a main correspondence effect (31 ms) was found, $F(1,12) = 19.5$, $p = .001$, which opposed to the condition with visual cues, was unaffected by cue, $F(1,12) = 0.03$.

ANOVAs for *auditory targets* revealed main effects of block type, $F(1,12) = 160.1$, $p < .001$, cue modality, $F(1,12) = 11.8$, $p = .005$, and correspondence, $F(1,12) = 56.0$, $p < .001$. Significant interactions were

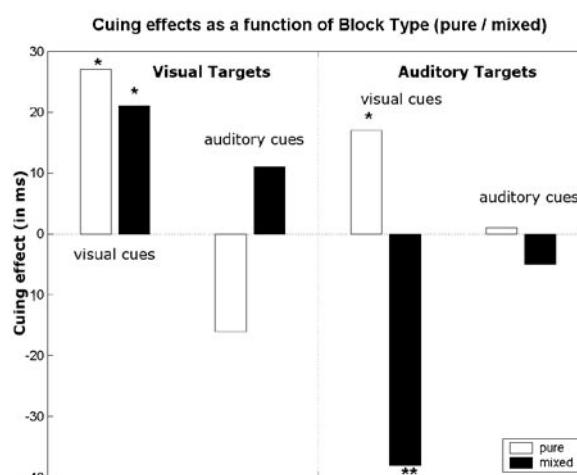


Figure 2.

*Cuing effects for visual and auditory targets preceded by visual or auditory cues in pure blocks (only one target modality) and mixed blocks (both visual and auditory targets). Cue modality varied from trial to trial, and was consequently always intermixed. Significant cuing effects are indicated with * ($p < 0.05$) or ** ($p < 0.005$).*

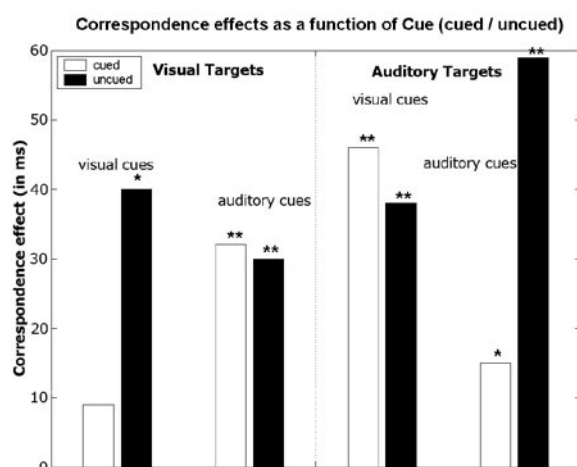


Figure 3.

*Correspondence effects for visual and auditory targets preceded by visual or auditory cues (averaged across block type) when targets occurred at cued or uncued locations. Significant correspondence effects are indicated with * ($p < 0.05$) or ** ($p < 0.005$).*

found between block type and cue, $F(1,12) = 20.8$, $p = .001$, between block type, cue, and cue modality, $F(1,12) = 10.4$, $p = .007$, and between cue modality, cue, and correspondence, $F(1,12) = 9.1$, $p = .011$. No interactions were observed involving both the factors correspondence and block type, $F(1,12) < 2.1$.

Further decomposition for visually cued auditory targets revealed a main effect of block type, $F(1,12) = 102.4$, $p < .001$, and correspondence (42 ms), $F(1,12) = 36.9$, $p < .001$, and a significant interaction between cue and block type, $F(1,12) = 44.0$, $p < .001$, which reflects opposed cuing effects in pure and mixed blocks

(see Fig. 2) of -37 ms (slowing for cued as compared to uncued targets) and 18 ms.

In case of auditory targets preceded by auditory cues, a main effect of block type, $F(1,12) = 202.7$, $p < .001$, and correspondence was found, $F(1,12) = 36.3$, $p < .001$. The correspondence effect interacted with block type, $F(1,12) = 9.3$, $p = .01$, being larger in pure blocks than in mixed blocks (48 vs. 26 ms). Finally, the correspondence effect interacted with cue, $F(1,12) = 13.4$, $p = .003$, being smaller for cued targets than for uncued targets (16 vs. 58 ms, see Fig. 3).

PC. Mean PCs are presented in Table 1 and Table 2. ANOVAs revealed a nearly significant effect of target modality, $F(1,12) = 4.2$, $p = .063$, an effect of cue modality, $F(1,12) = 5.9$, $p = .032$, and a significant interaction between block type, target modality and cue modality, $F(1,12) = 4.8$, $p = .05$. A main correspondence effect was found, $F(1,12) = 11.5$, $p = .005$, with more accurate responses for corresponding as compared to noncorresponding targets (92.7 % vs. 90.8 %). An interaction between block type, target modality and correspondence, $F(1,12) = 4.8$, $p = .048$, and between block type, target modality, correspondence and cue was found.

Separate analyses per target modality revealed only a trend to a correspondence effect in case of visual targets, $F(1,12) = 4.6$, $p = .053$, and in case of auditory targets, we observed an effect of cue modality, $F(1,12) = 5.3$, $p = .04$, a trend to a correspondence effect, $F(1,12) = 4.4$, $p = .057$, and a trend to an interaction between correspondence, cue, and block type.

Response force. An analysis on peak amplitude of the exerted response force with the factors block, target modality, cue modality, cue, and correspondence revealed a main effect of target modality, $F(1,12) = 5.1$, $p = .044$, with slightly more forceful responses for visual than for auditory targets (4.8 vs. 4.6 N), a nonsignificant trend to an effect of correspondence, $F(1,12) = 3.9$, $p = .073$, (4.74 vs. 4.67 N), but no other effects, $F(1,12) < 3.1$, $p > .10$.

A further exploration per target modality revealed a significant correspondence effect in case of visual targets, $F(1,12) = 5.7$, $p = .035$, with slightly more forceful responses on corresponding than on noncorresponding trials (4.85 vs. 4.76 N), whereas no effects were found for auditory targets, $F_s < 2.9$.

DISCUSSION

In the current study, we examined whether an extended version of the contingent attentional capture

account by Folk et al. (1992), can elucidate cuing effects in multimodal settings with visual and auditory stimuli. We were especially interested to what extent cuing effects depend on the presence of only one target modality in so-called "pure" blocks as compared to the presence of two target modalities in "mixed" blocks.

Evaluation of the extended contingent attentional capture account

In line with the contingent attentional capture account, we reasoned that an extended version for multimodal settings would propose that capture depends on top-down factors. In three conditions, auditory and visual exogenous cues preceded targets being either solely visual or auditory in pure blocks or equally probable visual or auditory in mixed blocks. Cuing effects of visual exogenous cues were predicted to be largest in pure visual blocks, intermediate in mixed blocks and smallest in pure auditory blocks. For auditory exogenous cues, an opposite pattern was expected to occur. Our findings (see Fig. 2) with visual cues were not in line with the proposed hypothesis. Clear positive cuing effects were found on RT to visual targets in pure and mixed target blocks (27 vs. 21 ms), but no difference was found between these blocks. In case of auditory targets, visual cues also resulted in a positive cuing effect in pure blocks (17 ms), but a strong negative cuing effect was found in mixed blocks (-38 ms). Obviously, these data indicate that cuing effects of visual onsets are context-dependent, but they are not in accordance with our predictions based on the extended contingent attentional capture account. With regard to auditory cues, a trend to a cuing effect was found in mixed blocks with visual targets, and a trend to an opposite effect was found in a pure block with visual targets. However, the dependency of the cuing effect on block type proved to be significant, which indicates that effects of auditory onsets are also context-dependent.

The tendency to an opposite cuing effect of auditory cues in pure blocks with visual targets is in contrast with other findings in our laboratory with the same equipment and nearly identical stimulus locations. In those studies, clear positive cuing effects were found when auditory cues preceded to be discriminated visual targets (Santangelo et al., 2006; Van der Lubbe & Postma, 2005). The only major difference between these studies and the current condition is the mixing of visual and auditory onsets within a block of trials. One could argue that mixing these onsets results in other

top-down settings, although effects may also be caused by interference from previous trials (i.e. sequential effects). To examine the latter possibility, we performed an analysis with cue modality on the preceding trial as additional factor, but found no support for an explanation of effects in terms of sequential effects⁵. As a consequence, the discrepancy is probably due to a change in top-down settings when cue modality varies from trial to trial. Specifically, as auditory cues are known to be irrelevant whereas visual stimuli are possibly relevant, there may be a tendency to ignore the locus of auditory stimuli, resulting in the observed interaction. A somewhat related view was forwarded by Ward, McDonald and Lin (2000) who argued that increased complexity of the environment (number of cue modalities, target modalities, etc.), may be a crucial factor, with diminished processing of the location of auditory cues in more complex environments. Thus, a complex environment (e.g. mixing cue modalities) might nullify the effects of auditory cues.

Another aspect of the results that deserves our attention is the absence of cuing effects when auditory targets are preceded by auditory cues, irrespective of block type. As space is not so likely to play an important role for pitch discrimination, this result may appear not so surprising (see Spence & Driver, 1994). Nevertheless, spatial information provided by the visual cues appears to play an important role, indicating that the relevance of space may lie within another domain, for example, the selection of action-relevant locations at a supramodal level (e.g. see Van der Lubbe, Neggers, Verleger, & Kenemans, 2006). In line with this, the locus of the auditory stimuli indeed affected the speed of responses, which will be focused upon further in the section on the Simon effect.

Irrespective of the cued location, inspection of Table 1 additionally reveals that responses to auditory targets were very slow in mixed blocks. A likely reason is that keeping the appropriate S-R mapping active will be more difficult for auditory targets as high or low pitches are more relative and difficult to memorize than triangles pointing up- or downwards. Of course, this memory factor cannot account for the context-dependent cuing effects. Finally, one might raise the question whether observed effects are possibly due to speed-accuracy tradeoff and differences in arousal. However, no relevant effects were found on response accuracy (see Table 1) and response force, which indicates that these factors played no important role. We only observed a just significant effect of more forceful responses after visual than after auditory targets, which suggests that opposed to our

suggestion visual rather than auditory targets were slightly more arousing.

How then can we explain the current data? First, we will focus on a modified version of the contingent attentional capture account forwarded by Ansorge and Heumann (2005; see Theeuwes, Atchley & Kramer, 2000). Ansorge and Heumann argued on the basis of electrophysiological data that onsets automatically attract attention, but that the speed of attention withdrawal may be dependent on top-down settings (for comparable hypotheses to explain other task-dependent cuing effects see Klein, 2000; Lupiáñez, Miliken, Solano, Weaver & Tipper, 2001; Van der Lubbe et al., 2005, Van der Lubbe et al., 2006). Does an extended version of this hypothesis to multimodal settings provide an explanation for the observed effects in our study? In case of to be discriminated visual targets in pure visual blocks, top-down settings should be set to visual stimuli as no other imperative stimuli are presented. As a consequence, when auditory cues occur then attention may already be rapidly withdrawn from the cued location, resulting in a reduced or even negative cuing effect. However, when top-down settings are set to visual and auditory stimuli in mixed blocks then an opposite effect may be found for auditory cues as attention will not be rapidly withdrawn from the auditory cued location. The pattern of results obtained with visual targets (left panel of Fig.2) can indeed be explained along these lines. However, application of the same logic to auditory targets should result in a positive cuing effect for visual cues in mixed blocks and a smaller or even negative cuing effect in pure blocks. Our results evidently show an opposite pattern (left part of right panel). As a consequence, this proposal cannot account for our results.

In a very recent study of Mazza et al. (2006) some comparable results were found with a version of the orthogonal cuing paradigm of Spence and Driver (1997a)⁶. In their first experiment, visual and auditory targets presented to the left or right were preceded by nonpredictive visual or auditory cues with either a short or a long SOA (150/700 ms). The location and modality of cues and targets varied randomly from trial to trial, hence, the trials with short SOAs in this experiment resemble our mixed target blocks. In case of visual targets and short SOAs, positive cuing effects were found with visual and auditory cues, and in case of auditory targets, a reversed non-significant tendency was present with visual cues and a positive cuing effect was found with auditory cues. Thus, in line with our findings, their data pattern suggested opposite cuing effects

with visual cues for auditory as compared to visual targets. Mazza et al. made some suggestions that effects of exogenous cues might be specific depending on top-down settings to target modality, which indeed appeared to be the case in our own study. Quite relevant are the findings by Turatto, Benso, Galfano and Umiltà (2002, see also Turatto, Galfano, Bridgeman & Umiltà, 2004; Spence & Driver, 1997b). First, regardless of any expectancy, there appears to be a general cost in responding to targets when stimulus modality changes, which was denoted as the modality shift effect (MSE). Turatto et al. (2002) additionally observed that in pure blocks with to be detected visual targets, modality of a previous warning stimulus had no effect, whereas it had an effect in pure blocks with auditory targets as detection of an auditory target was slowed down when preceded by a visual warning stimulus. Thus, visual inputs tend to dominate over auditory inputs ("visual dominance"), which may be due to a higher priority for visual inputs⁷. In another recent study by Rodway (2005) the focus was also on the difference between auditory and visual warning cues preceding targets. In this study, it was additionally suggested that it may be harder to disengage attention from a visually cued location than from an auditory cued location whereas at the same time visual cues may be more effective due to their higher spatial resolution. Possibly, the aforementioned factors (MSE, and visual dominance, possibly related to differences in the disengagement of attention and spatial resolution) are responsible for our findings. Namely, after a visual cue there will be priority for visual stimuli, due to the absence of a modality shift, leading to an advantage for visual targets in pure and mixed blocks, and leading to inattention for auditory targets on the cued location (in mixed target blocks) unless top-down settings are set to auditory targets. Of course, this explanation is rather adhoc, and additional experiments in which cue modality is either pure or mixed seem required to clarify what is going on. Another interesting idea is a partial replication of the experiment with visual and tactile stimuli as spatial resolution in the tactile modality will be higher than in the auditory modality.

Influence of exogenous cues on the Simon effect in multimodal settings

A secondary issue of the current study concerned the presence of a Simon or correspondence effect,

which has been considered as "a tendency to react towards the stimulus source" (e.g. Simon, 1990). Overall, a very clear Simon effect was found (33 ms), possibly due to the use of rather eccentric stimulus locations. Interestingly, the overall analyses revealed that the Simon effect interacted with cue, being smaller for cued targets than for uncued targets (see Fig.3). Moreover, effects additionally depended on cue and target modality. A separate analysis for visual targets showed that the Simon effect was modulated by cue location in case of visual cues, but not in case of auditory cues (see Fig.3). An opposite pattern was found for auditory targets, with no modulation of the Simon effect in case of visual cues, and a modulation in case of auditory cues. These findings seem rather interesting, and appear to have several implications. First of all, they demonstrate that the Simon effect is under certain conditions dependent on the current focus of attention, which contrast with the view that the Simon effect is solely due to the locus of a target stimulus with respect to a reference at fixation, the referential coding account (e.g. Hommel, 1993). In line with the current findings, a forthcoming paper (Abrahamse & Van der Lubbe, submitted) on the relation between spatial attention and the Simon effect, suggests that the Simon effect may be understood as due to the production of spatial codes related to attentional orienting (e.g. see Rubichi, Nicoletti, Iani & Umiltà, 1997), with the important addition that these spatial codes may be generated not only when selecting a target at a specific location, but also when selecting a location for other purposes, such as the execution of a hand- or eye-movement. Secondly, the observation of a large Simon effect for auditory targets demonstrates that space for the auditory modality is not irrelevant but may play a more important role for linking perception with action than for pitch perception. Third, spatial cues only reduced the Simon effect in unimodal settings, but not in crossmodal settings. These results might mean that a secondary step of attentional orienting may be necessary in crossmodal settings, leading to equivalent Simon effects for cued and uncued targets in these settings. A reason might be that the links between different spatial maps of the auditory and visual modality are not automatic but require an additional transformation, possibly induced at target onset, leading to new activation of a spatial code, with the consequence that the Simon effect has the same magnitude on cued and uncued trials in crossmodal settings. Clearly, more research on these issues seems required. Interesting manipu-

lations are again the influence of blocking cue modality and the employment of tactile stimuli.

In conclusion, our results revealed that cuing effects of auditory and visual onsets are not fully automatic as their orienting effects were context-dependent. Nevertheless, the observed findings were neither in line with an extended version of the contingent attentional capture account nor with an extension in terms of a top-down influence on the speed of attention withdrawal. Obviously, more work is required on multimodal settings to clarify aspects like visual dominance and the influence of top-down factors, and the origin and modulation of the Simon effect. These findings will have important implications for more realistic settings in everyday life, in which we are commonly faced with multimodal stimulation.

Acknowledgements

The current study was supported by a grant from The Netherlands Organization for Fundamental Research to Albert Postma (NWO: 440-20-000). Thanks are due to Michiel Reimerink for his help in measuring the participants, to Piotr Jaśkowski for his help in the force analyses, and to two reviewers for comments on an earlier draft of this paper.

Notes

¹ Exceptions have been observed when using a simple detection task, e.g., see Van der Lubbe, Vogel and Postma (2005), and Van der Lubbe, Havik, Bekker, and Postma (2006), which seem at least partially due to foreknowledge of the required response in these tasks. To avoid unnecessary complications, we will not focus on these tasks.

² Both types of covert orienting do of course not refer to the actual movement of our eyes towards specific locations (overt orienting), although overt orienting often follows covert orienting.

³ This conclusion is based on a null effect. Therefore, an advocate of the view that upholds the special status of abrupt onsets could argue that the statistical power of this experiment was too low.

⁴ We thank an anonymous reviewer of this paper for raising this interesting issue.

⁵ A separate analysis for visual targets with the additional factor cue modality on trial n-1 revealed no significant effects, $F(1,12) < 1.2$.

⁶ An important difference with the Posner paradigm employed in the current study is that cues and targets in the orthogonal cuing paradigm occur on the same side but not at the same location. As a consequence, application of the logic of Ansorge & Heumann (2003)

would suggest that no effects of these cues should be present as cue locations provide no information about target locations. A multitude of studies by Spence & Driver clearly revealed that these cues are effective, signifying that the contingent attentional capture hypothesis encounters problems in multimodal settings.

⁷ In the study by Turatto et al. memory factors play no important role as detection responses were required to auditory targets rather than pitch discrimination responses.

References

- Abrahamse, E.L., & Van der Lubbe, R.H.J. (submitted). Endogenous orienting modulates the Simon effect.
- Ansorge, U., & Heumann, M. (2003). Top-down contingencies in peripheral cuing: the roles of color and location. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 937-948.
- Ansorge, U., & Heumann, M. (2005). Shifts of visuospatial attention to invisible (metacounter-masked) singletons: Clues from reaction times and event-related potential. *Advances in Cognitive Psychology*, 2, 61-76.
- Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General*, 134, 207-221.
- 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.
- Folk, C.L., & Remington, R. (1999). Can new objects override attentional control settings? *Perception & Psychophysics*, 61, 727-739.
- Hommel, B. (1993). The role of attention for the Simon effect. *Psychological Research*, 55, 208-222.
- Hopfinger, J.B., & West, V.M. (2006). Interactions between endogenous and exogenous attention on cortical visual processing. *NeuroImage*, 31, 774-789.
- Ivanoff, J., Klein, R.M., & Lupiáñez, J. (2002). Inhibition of return interacts with the Simon effect: An omnibus analysis and its implications. *Perception & Psychophysics*, 64, 318-327.
- Jaśkowski, P., Van der Lubbe, R.H.J., Wauschkuhn, B., Wascher, E., & Verleger, R. (2000). The influence of time pressure and expectancy on response force in an S1-S2 paradigm. *Acta Psychologica*, 105, 89-105.
- Jaśkowski, P., & Włodarczyk, D. (2005). Task modulation of the effects of brightness on reaction

- time and response force. *International Journal of Psychophysiology*, 61, 98-112
- Klein, R.M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, 4, 138-146.
- Lupiáñez, J., Miliken, B., Solano, C., Weaver, B., & Tipper, S.P. (2001). On the strategic modulation of the time course of facilitation and inhibition of return. *The Quarterly Journal of Experimental Psychology*, 54A, 753-773.
- Mazza, V., Turatto, M., Rossi, M., & Umiltà, C. (2006). How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia*,
- McCormick, P.A. (1997). Orienting of attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 168-180.
- McDonald, J.J., & Ward, L.M. (2000). Involuntary listening aids seeing: evidence from human electrophysiology. *Psychological Science*, 11, 167-171.
- Posner, M.I., Cohen, Y. (1984). Components of visual orienting. In: Bouma H, Bouwhuis DG (eds). *Attention and performance X*. Erlbaum, Hillsdale, NJ, pp 531-555.
- Rodway, P. (2005). The modality shift effect and the effectiveness of warning signals in different modalities. *Acta Psychologica*, 120, 199-226.
- Rubichi, S., Nicoletti, R, Iani, C., & Umiltà, C. (1997). The Simon effect occurs relative to the direction of an attention shift. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1353-1364.
- Santangelo, V., Van der Lubbe, R.H.J., Olivetti Belardinelli, M., & Postma, A. (2006). Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: a comparison of reflexive orienting mechanisms. *Experimental Brain Research*, 173, 40-48.
- Schmitt, M., Postma A., & De Haan E. (2000). Interactions between exogenous auditory and visual spatial attention. *Quarterly Journal of Experimental Psychology*, 53, 105-130.
- Simon, J.R. (1990). The effects of an irrelevant directional cue on human information processing. In: Proctor, R.W, and Reeve, T.G. (Eds.). *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). Amsterdam: North Holland.
- Simons, D.J., & Levin, D.T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1, 261-267.
- Spence, C.J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 335-374.
- Spence, C.J., & Driver, J. (1997a). Audiovisual links in exogenous covert spatial attention. *Perception & Psychophysics*, 59, 1-122.
- Spence, C.J., & Driver, J. (1997b). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, 59, 384-403.
- Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, 49, 83-90.
- Theeuwes, J., Atchley, P., & Kramer, A.F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII*, (pp 105-125), Cambridge, MA, MIT Press.
- 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., Kramer, A.F., Hahn, S., & Irwin, D.E., & Zelinsky, G.J. (1999). Influence of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1595-1608.
- Turatto, M., Benso, F., Galfano, G., & Umiltà, C. (2002). Nonspatial attentional shifts between audition and vision. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 628-639.
- Turatto, M., Galfano, G., Bridgeman, B., & Umiltà, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Experimental Brain Research*, 155, 301-310.
- Van der Heijden, A.H.C. (1992). *Selective attention in vision*. Routledge: London, New York.
- Van der Heijden, A.H.C. (2004). *Attention in vision. Perception, communication and action*. Psychology Press: Hove, New York.
- Van der Lubbe, R.H.J., Havik, M.M., Bekker, E.M., & Postma, A. (2006). Task-dependent exogenous cuing effects depend on cue modality. *Psychophysiology*, 43, 145-160.
- Van der Lubbe, R.H.J., Keuss, P.J.G., & Stoffels, E.-J. (1996). Threefold effect of peripheral precues: alertness, orienting, and response tendencies. *Acta Psychologica*, 94, 107-118.
- Van der Lubbe, R.H.J., Neggers, S.F.W., Verleger, R., & Kenemans, J.L. (2006). Spatiotemporal overlap between brain activation related to saccade preparation and attentional orienting. *Brain Research*, 1072, 133-152.
- Van der Lubbe, R.H.J., & Postma, A. (2005). Interruption from irrelevant auditory and visual onsets even when attention is in a focused state. *Experimental Brain Research*, 164, 464-471.

- Van der Lubbe, R.H.J., Vogel, R.O., & Postma, A. (2005). Different effects of exogenous cues in a visual detection and discrimination task. Delayed attention withdrawal and/or speeded motor inhibition? *Journal of Cognitive Neuroscience*, 17, 1829-1840.
- Van der Lubbe, R.H.J., & Woestenburg, J.C. (1999). The influence of peripheral cues on the tendency to react towards a lateral relevant stimulus with multiple-item arrays. *Biological Psychology*, 51, 1-21.
- Ward, L.M., McDonald, J.J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics*, 62, 1258-1264.
- Wu, S.-C., & Remington, R.W. (2003). Characteristics of covert and overt visual orienting: evidence from attentional and oculomotor capture. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 1050-1067.
- 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.