

# Multisensory top-down sets: Evidence for contingent crossmodal capture

Frank Mast<sup>1</sup> · Christian Frings<sup>1</sup> · Charles Spence<sup>2</sup>

Published online: 6 May 2015  
© The Psychonomic Society, Inc. 2015

**Abstract** Numerous studies that have investigated visual selective attention have demonstrated that a salient but task-irrelevant stimulus can involuntarily capture a participant's attention. Over the years, a lively debate has erupted concerning the impact of *contingent top-down control settings* on such stimulus-driven *attentional capture*. In the research reported here, we investigated whether top-down sets would also affect participants' performance in a multisensory task setting. A nonspatial compatibility task was used, in which the target and the distractor were always presented sequentially from the same spatial location. We manipulated target–distractor similarity by varying the visual and tactile features of the stimuli. Participants always responded to the visual target features (color); the tactile features were incorporated into the participants' top-down set only when the experimental context allowed for the tactile feature to be used in order to discriminate the target from the distractor. Larger compatibility effects after bimodal distractors were observed only when the participants were searching for a bimodal target and when tactile information was useful. Taken together, these results provide the first demonstration of nonspatial contingent crossmodal capture.

**Keywords** Contingent capture · Crossmodal attention · Multisensory integration · Action control · Touch · Vision

---

✉ Frank Mast  
mastfra@uni-trier.de

<sup>1</sup> Department of Psychology, University of Trier, Trier, Germany

<sup>2</sup> Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

An extensive body of empirical research on the topic of selective (visual) attention has revealed that the peripheral presentation of a cue stimulus automatically attracts spatial attention (for the original task, see Posner, 1980; but see also Jonides, 1981; Yantis & Jonides, 1984). The exogenous capture of spatial attention can be observed even though a cue may be entirely task-irrelevant in terms of predicting the location (or identity) of the subsequent target. The claim that has often been made in the literature is that the presentation of a peripheral cue *captures attention* in a purely stimulus-driven manner, and that the underlying processes are independent of the participant's current behavioral goals (see Theeuwes, 2010, for a review). Bottom-up *attentional capture* is thought to be modulated by the saliency of the cue, based on the assumption that those stimuli that happen to stand out from the background in one or more feature dimensions are more salient than the background, and hence have the potential to exogenously attract a participant's spatial attention. The ability of various feature dimensions to automatically attract attention has been investigated by a number of researchers over the years. Researchers have, for instance, observed that attention can be captured by changes in various stimulus dimensions, such as an abrupt stimulus onset (Theeuwes, 1991; Yantis & Jonides, 1984), a change in stimulus luminance (Rauschenberger, 2003), or a change in the color or shape of a stimulus (Theeuwes, 1992).

There has been much debate about the impact of top-down mechanisms on the ability of specific stimuli to exogenously attract attention (see Burnham, 2007, and Theeuwes, 2010, for reviews). One of the most influential hypotheses concerning the impact of top-down mechanisms is the *contingent capture hypothesis* originally put forward by Folk, Remington, and Johnston (1992; see also Gibson & Kelsey, 1998, for a related notion). According to this hypothesis, participants create top-down sets in order to rapidly localize and identify the expected target stimuli. Those perceptual features that are associated

with the appearance of the target stimulus are incorporated within these top-down sets (e.g., the abrupt onset of a stimulus, a certain color or shape). In order to capture attention, a distractor needs to share the critical feature(s) with the top-down set. In their original study, Folk et al. (1992) had their participants identify either a red target stimulus (a color singleton) presented among three white distractors, or else respond to the appearance of a single target stimulus (an onset singleton). The target display was preceded by either of two displays, a color singleton or an onset singleton. Intriguingly, only those cues that matched the relevant target feature gave rise to an automatic and involuntary shift of the participant's spatial attention toward the cued location (i.e., color cues only affected the participants' responses to the color targets, but not to the onset targets, and vice versa). If the subsequent target happened to have been presented from the same spatial location as the cue stimulus, processing was speeded up. These findings have now been replicated in numerous studies involving both spatial (Anderson & Folk, 2010; Ansorge & Heumann, 2003, 2004; Eimer, Kiss, Press, & Sauter, 2009; Folk & Remington, 1998; Folk, Remington, & Wright, 1994) and nonspatial *attentional blink* (AB) tasks (Folk, Leber, & Egeth, 2008; see also Folk et al., 2002, for another AB task, but with spatial distractors). Taken together, such results provide robust support for the importance of top-down mechanisms in attentional capture. According to the contingent capture hypothesis, participants integrate those features into their top-down sets that are relevant for the rapid localization and identification of the target. Whether or not a feature is relevant depends on whether or not its appearance is correlated with the appearance of the target stimulus.

### Crossmodal attention and contingent capture

To date, the debate concerning the impact of top-down sets on attentional capture has been based on unimodal studies—that is, input from only a single sensory modality has been involved (see van der Lubbe & Van der Helde, 2006, for an exception—although these authors used an audiovisual task, all of the effects that were observed could be attributed to a supramodal spatial attention; see Spence, 2013, for a review of crossmodal spatial attention; see Spence, 2010, for a review of studies challenging the automaticity of crossmodal spatial-cuing effects). The present study was designed to investigate whether top-down sets are as important for the guidance of crossmodal attention as they are for the guidance of visual selective attention (see Shore & Simic, 2005, for a study of top-down influences on visuotactile integration, resulting from variations in the proportion of congruent versus incongruent stimuli in a spatial visuotactile congruency task). In the present study, a variant of the response-priming task (see Mast & Frings, 2014) with two sequentially presented stimuli, both

presented at a clearly suprathreshold level, was used (see Neumann & Klotz, 1994; for a review of the literature on subliminal priming, see Kiesel, Kunde, & Hoffmann, 2007; see also Van den Bussche, Van den Noortgate, & Reynvoet, 2009, for a recent meta-analysis).

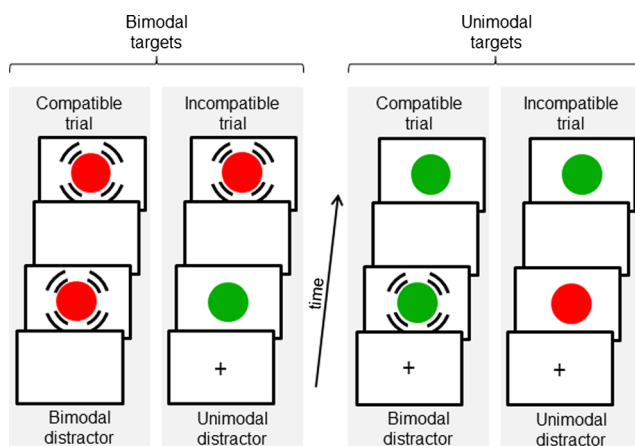
A nonspatial priming task was used instead of an exogenous cuing task because the utilization of a spatial task might have been problematic, due to a supramodal spatial feature (e.g., the tactile stimulation captured attention because it appeared from one of the possible visual target locations; cf. Spence, 2013). Due to the fact that both stimuli (the prime and the subsequently presented target) were visible to the participants (note that in a typical response-priming experiment, the participants are normally not aware of the prime), the instruction was given to try to ignore the first stimulus and to respond as rapidly as possible to the second. Therefore, the term *distractor* (rather than *prime* or *cue*) will be used here, because the participants were explicitly informed that they should ignore the first stimulus (which interfered with responding on 50 % of all trials) and exclusively respond to the second stimulus (i.e., the target). During each trial, the target and distractor indicated a certain response; in the compatible trials, both of the stimuli were mapped onto the same response, whereas in the incompatible trials they were mapped onto different responses. In the compatible trials, reaction times to the target are normally speeded while error rates tend to decrease, relative to incompatible trials.

In the present study, the supraliminal variant of the response-priming paradigm was extended to a multisensory setting by adding additional tactile stimulation (see Wesslein, Spence, & Frings, 2014, for a review of tactile distractor processing). The task itself was visual, the target (either a red or a green circle) and the distractor (either a red or a green circle) were both presented successively from the same location. On the basis of studies by Ansorge and Heumann (2003, 2004), the similarity between the target and the distractor was manipulated. In their study, the participants were asked to search for a target that was indicated by a specific color (e.g., red) that could be presented from one of two possible locations. The target stimuli were preceded by an unpredictable spatial cue that was either similar to the target (e.g., yellowish red) or dissimilar to it (e.g., bluish green). The size of the cuing effect was found to vary as a function of the target–cue similarity; larger cuing effects were observed for distractors that were similar to the target than for target-dissimilar distractors. Recently, Mast and Frings (2014) further demonstrated that, according to the task requirements, top-down sets could be compiled of multiple feature dimensions (e.g., location and color; see Awh, Belopolsky, & Theeuwes, 2012, for a review). The strength of attentional capture by a distractor was found to vary as a function of the overlap between the features of the distractor and the features incorporated into the participants' top-down sets. On the basis

of these studies, our prediction was that participants would compile their top-down sets from multiple features, and what is more, from features or stimuli presented in different sensory modalities.

The present study was conducted in order to investigate the postulated compilation of multisensory top-down sets. Thus, instead of manipulating target–distractor similarity by varying one of the target’s visual features, we manipulated it by the addition of a tactile feature instead. By presenting either unimodal or bimodal targets, we attempted to induce two different top-down sets (unimodal vs. bimodal). In Experiment 1a, the target modality was manipulated on a between-subjects basis. The participants in the bimodal target condition were always stimulated by an additional tactile stimulus during the presentation of the visual target, whereas the participants in the unimodal target condition never received such tactile stimulation during the presentation of the target (see Fig. 1). Relative to the purely visual target condition, the constant co-occurrence of the target with the tactile stimulus (the bimodal target condition) should have led to the adaption of the participants’ top-down sets. Those participants who always experienced the visual target stimulus in the absence of any tactile stimulation should presumably have based their top-down sets on purely visual features, whereas those participants in the bimodal target condition should have added an additional tactile feature to their top-down sets. Note that the tactile stimulation was not linked to either of the two response alternatives.

In line with the assumption of the extended, bimodal top-down set, target–distractor similarity was manipulated by means of a tactile stimulation during the presentation of the distractor. The central prediction here was that in the bimodal

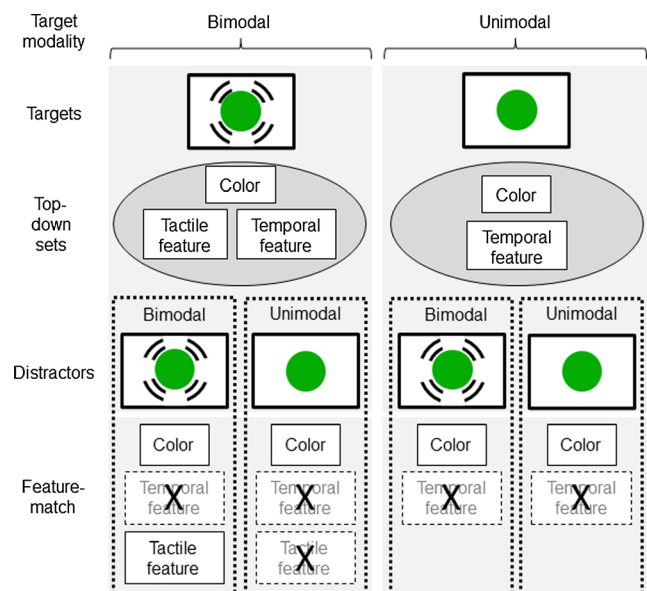


**Fig. 1** Sequence of events for trials in the bimodal (left side) and unimodal (right side) target conditions. In the compatible trials, the target and distractor were both presented in the same color, whereas in the incompatible trials they were presented in different colors. Throughout the bimodal target condition, the target was always accompanied by a vibrotactile stimulus (indicated by the shock waves around the stimuli); in the unimodal target condition, the target was never presented with a vibrotactile stimulus.

target condition, bimodal distractors would attract more attention and more processing resources than unimodal distractors, due to their increased similarity to the target. As a consequence, it was assumed that bimodal distractors would give rise to larger compatibility effects than would unimodal distractors. In the unimodal target condition, however, no such difference was predicted, since the top-down sets should be compiled from visual and temporal features; consequently, tactile stimulation should neither increase nor decrease the similarity between the visual target and the visual distractor (see Fig. 2 for a graphical explanation).

## Experiment 1a

In Experiment 1a, the targets and distractors were either unimodal or bimodal. The target modality was manipulated between subjects (i.e., half of the participants received only bimodal targets, whereas the other half received only unimodal targets). In order to prevent the participants from allocating their attention to a specific point in time (see Nobre, 2001; Nobre & Coull, 2010), the stimulus onset asynchrony (SOA) between the appearances of the distractor and the target was varied randomly on a trial-by-trial basis (120, 160, or 180 ms). In addition, the impact of the distractors’ predictability was investigated. Therefore, two different modes of stimulus presentation were utilized (cf. Mast,



**Fig. 2** The figure depicts the logic of why a difference in the sizes of the compatibility effects for bimodal and unimodal distractors should be found, but only in the bimodal target condition and not in the unimodal target condition. Only when the tactile stimulation is presented together with the visual target should the tactile stimulation be implemented into the participants’ top-down sets. As a consequence, tactile stimulation during distractor presentation would increase the feature match between the distractor and the top-down sets only for bimodal targets. Finally, the feature match should predict the size of the resulting compatibility effects. See the text for further information.

Frings, & Spence, 2014): The tactile stimulation during the presentation of the target was either totally predictable (fixed mode of presentation) or unpredictable (variable mode of presentation).

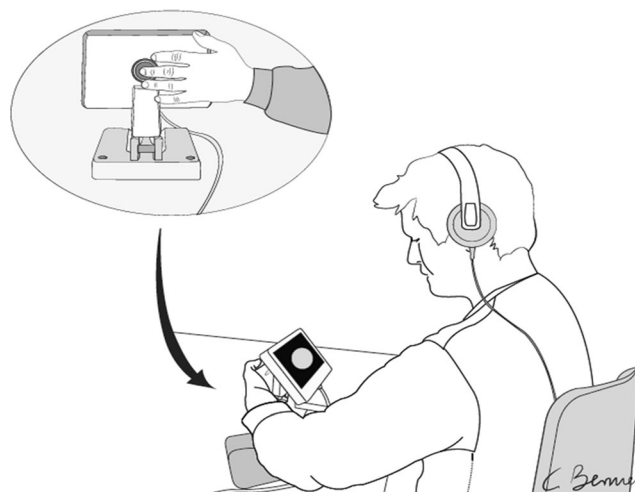
## Method

**Participants** A total of 30 students (26 women, four men; mean age 22 years) from the University of Trier took part in the study; 14 students participated in the bimodal target condition, and another 16 took part in the unimodal target condition. All of the participants reported normal or corrected-to-normal vision, and none of them reported any somatosensory impairment.

**Design** The participants were tested in a  $2 \times 2 \times 2 \times 2$  factorial design, with the three within-subjects factors of Response Compatibility (compatible vs. incompatible), Distractor Modality (unimodal vs. bimodal), and Distractor Presentation Mode (fixed vs. variable), and Target Modality (unimodal vs. bimodal) as a between-subjects factor. The participants were randomly assigned to the target modality conditions.

**Stimuli and apparatus** In order to reduce the amount of background environmental noise to a minimum, the experiment was conducted in a completely soundproofed room. The laboratory as well as the furniture were painted black, and all sources of illumination (e.g., from technical equipment) were eliminated. The instructions as well as the visual stimuli were presented on a 7-in. monitor (Model FT0070TM, Faytech Ltd., Henzen, China). The refresh rate of the monitor was 60 Hz, and it was placed approximately 12 cm in front of the participant's body midline. The participant's responses were detected with a standard PC mouse connected via a USB 2.0 port. The vibrotactile stimulus was presented by means of a tactor (Model C-2, Engineering Acoustic, Inc.) attached to the rear of the screen (see Fig. 3 for a schematic illustration of the experimental setup). The tactor was 3 cm in diameter and 0.8 cm thick, and the optimal frequency for its operation was approximately 250 Hz. White noise was presented over headphones in order to exclude any impact of the sounds caused by the operation of the skin transducers. When asked, the participants reported that they had not heard any sound associated with the tactile stimulation.

The visual stimuli were either green (CIE  $L^*a^*b^*$  values: 46, -52, 50) or red (CIE  $L^*a^*b^*$  values: 53, 80, 67) circles with a diameter of approximately  $1.72^\circ$ . The targets and distractors were always presented from the same central location on the screen, which was indicated by a fixation cross at the start of each trial. The tactile stimulus consisted of a



**Fig. 3** Schematic illustration of the experimental setting utilized in both experiments. Note that the tactile and visual stimuli were presented from the same direction with respect to the participant (though from different distances).

vibrotactile pulse, presented for the same duration as the visual stimulus to the participant's middle finger.

**Procedure** During each trial, two stimuli were presented successively: The first was the distractor and the second, the target. The participants were instructed to try to ignore the distractor and to respond as rapidly and accurately as possible to the identity of the visual target according to the stimulus response mapping that they had learned. Each trial started with the central presentation of the plus sign for 600 ms. This was immediately followed by the appearance of the distractor for 33 ms. Between the distractor and target displays was a random interval (the distractor–target SOA; 133, 167, or 200 ms) with a blank screen. Finally, the target appeared for 33 ms. The participants had to respond to the identity of the target (i.e., its color) within 2,030 ms of the onset of the target. After having responded to the color of the target, a further empty interval of 600 ms followed, before the start of the next trial. The presentation of the visual target stimulus was always accompanied by a tactile stimulation for the participants in the bimodal-target group, whereas for those in the unimodal-target group, the visual target was never presented together with tactile stimulation (see Fig. 1 for the different trial types).

Overall, the participants worked their way through four blocks of experimental trials. In each block, compatible and incompatible trials were intermixed randomly. In two of these experimental blocks (fixed blocks), the participants were informed that the distractor would either always (or never) be accompanied by tactile stimulation. Throughout these two experimental blocks, the participants could perfectly well predict the sequence of tactile events. During the two remaining



experimental blocks, the presence or absence of vibrotactile stimulation during the presentation of the distractor was unpredictable (variable blocks).

At the beginning of the experiment, the participants had to work their way through two short training phases. In the first training block, only one stimulus was presented, in order to facilitate the participants' learning of the stimulus–response mapping (16 trials). During the second training phase, two stimuli were always presented, the distractor and the target (48 trials). Feedback to the participants after each of their responses was provided in both of the training phases. Each of the four experimental blocks comprised 168 trials (84 compatible, 84 incompatible distractor–target sequences; SOA was orthogonally varied) and was preceded by 16 additional, warm-up trials. After every 40th trial, the participants were offered a break. Should the participant make three errors in a row, then he or she was offered another break.

## Results

Only those trials in which the participant responded correctly to the target were considered. Additionally, all trials in which the reaction time (RT) was shorter than 200 ms, as well as those trials with an RT that was 1.5 interquartile ranges above the third quartile of each participant's individual RT distribution (Tukey, 1977), were excluded from the data analyses. In all, 8.5 % of the trials were excluded from the analysis due to these restrictions. The mean RTs and error rates are highlighted in Table 1.

**RTs** A 2 (Response Compatibility: compatible vs. incompatible)  $\times$  2 (Distractor Modality: bimodal vs. unimodal)  $\times$  2 (Mode of Presentation: fixed vs. variable)  $\times$  2 (Target Modality: bimodal vs. unimodal) multivariate analysis of variance (MANOVA) with Pillai's trace as the criterion was conducted, with mean RTs as the dependent variable. The MANOVA revealed a significant main effect of response compatibility,  $F(1, 28) = 144.761, p < .001, \eta_p^2 = .838$ . That is, the participants responded more rapidly when the visual target and distractor were both mapped to the same response than when they were mapped to different responses. The main effect of distractor modality was also significant,  $F(1, 28) = 26.182, p < .001, \eta_p^2 = .483$ . The participants responded significantly more rapidly when the target was preceded by a bimodal distractor than when it was preceded by a unimodal distractor. The Response Compatibility  $\times$  Distractor Modality  $\times$  Mode of Presentation interaction was also significant,  $F(1, 28) = 6.557, p < .016, \eta_p^2 = .19$ . The sizes of the compatibility effects for the two distractor types (bimodal vs. unimodal) varied as a function of whether they were delivered in the fixed or the variable mode of presentation. Most importantly, however, the three-way interaction was further specified by

the target type for the two experimental groups,  $F(1, 28) = 2.932, p < .049$  (one-tailed),  $\eta_p^2 = .095$ . This result indicates that the sizes of the compatibility effects for unimodal and bimodal distractors varied as a function of the target modality and the mode of presentation. In order to interpret this four-way interaction, further analyses were conducted by separating the data from the two modes of presentation (fixed and variable) and utilizing the compatibility effects as the dependent variable.<sup>1</sup>

A 2 (Distractor Modality: bimodal vs. unimodal)  $\times$  2 (Target Modality: bimodal vs. unimodal) MANOVA with compatibility effects as the dependent variable and Pillai's trace as the criterion was conducted for the variable mode of presentation. The analysis did not reveal any significant main effect of target modality,  $F(1, 28) < 1$ , thus showing that the sizes of the compatibility effects did not differ between the two groups. The main effect of distractor modality reached significance,  $F(1, 28) = 6.537, p = .016, \eta_p^2 = .189$ . Larger compatibility effects were observed in those trials on which bimodal distractors were presented than in the trials with unimodal distractors. Most importantly, the Distractor Modality  $\times$  Target Modality interaction was also significant,  $F(1, 28) = 6.229, p = .019, \eta_p^2 = .182$ . Two additional one-sample *t* tests revealed that the sizes of the compatibility effects for bimodal ( $M = 145$  ms,  $SD = 53$  ms) and unimodal ( $M = 119$  ms,  $SD = 61$  ms) distractors only differed in the bimodal target condition,  $M_{\text{Diff}} = 26$  ms ( $SD = 30$  ms),  $t(13) = 3.219, p = .007$ . By contrast, in the unimodal target condition, no such difference in the sizes of the compatibility effects was observed,  $M_{\text{Diff}} = 1$  ms ( $SD = 26$  ms),  $t(15) = .048, p = .962$  (bimodal distractors,  $M = 132$  ms,  $SD = 63$  ms; unimodal distractors,  $M = 131$  ms,  $SD = 68$  ms). Thus, larger compatibility effects for bimodal than for unimodal distractors were observed in the bimodal target condition, but not in the unimodal target condition (see Fig. 4).

The same MANOVA was conducted for the fixed mode of stimulus presentation. Neither the main effect of distractor modality,  $F(1, 28) = 1.257, p = .272, \eta_p^2 = .043$ , or target modality (the between-subjects factor),  $F(1, 28) < 1$ , nor their interaction,  $F(1, 28) < 1$ , revealed a significant result. Thus, within the fixed mode of presentation, the size of the compatibility effect was not affected by the tactile stimulation during either distractor or target presentation.

**Error rates** Initially, the error rates were also analyzed by means of a 2 (Response Compatibility: compatible vs. incompatible)  $\times$  2 (Distractor Modality: bimodal vs. unimodal)  $\times$  2 (Mode of Presentation: fixed vs. variable)  $\times$  2 (Target

<sup>1</sup> Note that the main effect of Target Modality, the between-subjects factor, did not reach significance,  $F(1, 28) < 1$ . Accordingly, the presented results are not confounded by a systematic difference between the groups.

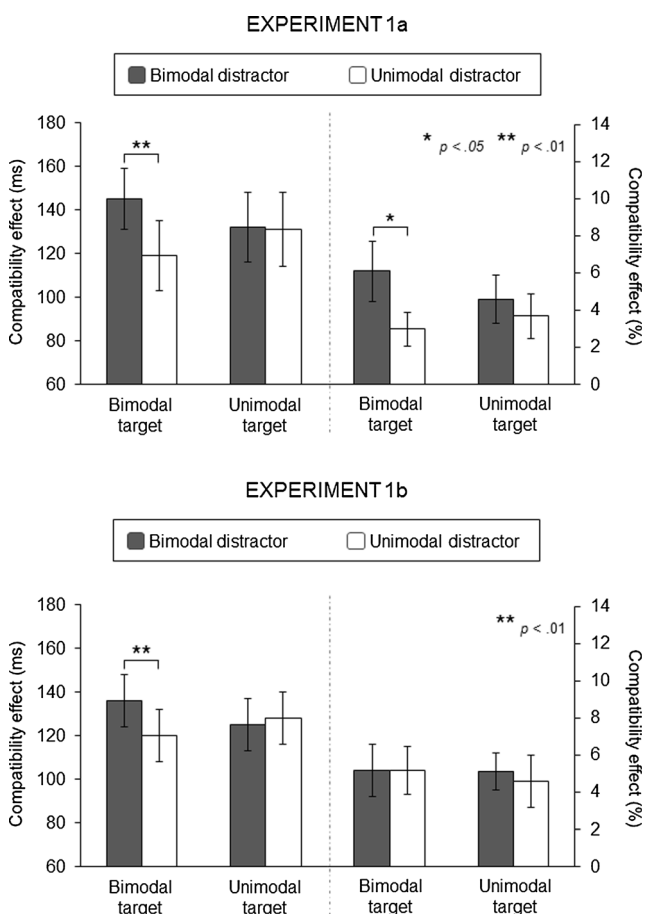
**Table 1** Mean reaction times (RTs, in milliseconds) and mean error rates (as percentages, in parentheses) as a function of response compatibility (incompatible vs. compatible), mode of presentation

Mode of Presentation		Bimodal Target		Unimodal Target	
		Bimodal Distractor	Unimodal Distractor	Bimodal Distractor	Unimodal Distractor
Variable	Incompatible	534 (6.9)	541 (4.1)	552 (6.1)	556 (4.8)
	Compatible	389 (0.8)	421 (1.1)	420 (1.5)	425 (1.1)
	Comp. effect	145 (6.1)	120 (3.0)	132 (4.6)	131 (3.7)
Fixed	Incompatible	520 (4.7)	554 (5.5)	537 (5.7)	564 (3.9)
	Compatible	397 (1.4)	420 (1.0)	416 (1.3)	436 (0.6)
	Comp. effect	123 (3.2)	134 (4.6)	121 (4.5)	128 (3.3)

Compatibility effects (Comp. effect) were computed as the difference between response-incompatible and response-compatible trials.

Modality: bimodal vs. unimodal) MANOVA with the mean error rates as the dependent variable and with Pillai’s trace as the criterion. The main effect of response compatibility was

(variable vs. fixed), target modality (bimodal vs. unimodal), and distractor modality (bimodal vs. unimodal) in Experiment 1a



**Fig. 4** Reaction time (RT, in milliseconds; on the left side) and error rate (as percentages; on the right side) compatibility effects in Experiments 1a and 1b as a function of the target and distractor modalities. Compatibility effects were computed as the difference between response-incompatible and response-compatible trials. The error bars depict the standard errors of the means. Note that for Experiment 1a, only the results for the variable mode of presentation are shown. In Experiment 1b, only the variable mode of presentation was presented.

significant,  $F(1, 28) = 34.18, p < .001, \eta_p^2 = .550$ , with participants making fewer errors in the compatible than in the incompatible trials. The main effect of distractor modality once again was also significant,  $F(1, 28) = 11.61, p = .002, \eta_p^2 = .293$ . The participants made more errors after bimodal than after unimodal distractors. Additionally, the three-way Response Compatibility  $\times$  Distractor Modality  $\times$  Mode of Presentation interaction was significant,  $F(1, 28) = 6.43, p = .017, \eta_p^2 = .187$ . This interaction was further specified by the target modality. In other words, the four-way interaction was also significant,  $F(1, 28) = 9.00, p = .006, \eta_p^2 = .243$ . The same data separation as for the RT data was conducted for the error rates, and two more MANOVAs were conducted for the two different modes of presentation.<sup>2</sup>

A 2 (Distractor Modality: bimodal vs. unimodal)  $\times$  2 (Target Modality: bimodal vs. unimodal) MANOVA with compatibility effects as the dependent variable and Pillai’s trace as the criterion was conducted for the variable mode of presentation. Once again, the main effect of distractor modality reached significance,  $F(1, 28) = 8.613, p = .007, \eta_p^2 = .235$ . This means that larger compatibility effects were observed following bimodal than following unimodal distractors. Neither the main effect of target modality,  $F(1, 28) < 1$ , nor the interaction reached significance,  $F(1, 28) = 2.791, p = .106, \eta_p^2 = .091$ . In order to make the analysis of error rates comparable to that conducted with RTs, two additional one-sample *t* tests for the unimodal and bimodal target conditions were conducted. They revealed a significant difference in the sizes of the compatibility effects between the bimodal ( $M = 6.1\%$ ,  $SD = 6.0\%$ ) and the unimodal ( $M = 3.0\%$ ,  $SD = 3.4\%$ ) distractors, but only for the bimodal target condition,  $M_{Diff} = 3.1\%$  ( $SD = 4.0\%$ ),  $t(13) = 2.89, p = .013$ , and not for the unimodal target condition,  $M_{Diff} = 0.8\%$

<sup>2</sup> The main effect of the between-subjects factor Target Modality again did not reach significance,  $F(1, 28) < 1$ , so the mean error rates in the two experimental groups did not differ systematically.

( $SD = 3.3\%$ ),  $t(15) = 1.01$ ,  $p = .327$  (bimodal distractors,  $M = 4.6\%$ ,  $SD = 3.4\%$ ; unimodal distractors,  $M = 3.7\%$ ,  $SD = 4.3\%$ ) (see Fig. 4).

The analysis of the error rates in the fixed mode of presentation again revealed no significant results: Neither the main effect of distractor modality,  $F(1, 28) < 1$ , nor the mode of presentation,  $F(1, 28) < 1$ , nor their interaction,  $F(1, 28) = 3.536$ ,  $p = .071$ ,  $\eta_p^2 = .112$ , reached significance. The sizes of the compatibility effects in the fixed mode-of-presentation condition did not differ for the bimodal and unimodal distractors between the two groups.

## Discussion

In Experiment 1a, a supraliminal variant of the response-priming task was used with either unimodal targets (purely visual) or bimodal targets (visual stimuli that were accompanied by response-irrelevant tactile vibrations) for the two different groups of participants. In order to determine whether participants applied different top-down sets (unimodal vs. bimodal), the feature similarity between the targets and the distractors was manipulated. It was assumed that in the bimodal target group, the bimodal distractors would be more similar to the target than the unimodal distractors. By contrast, the presentation of an additional vibrotactile stimulus during the presentation of the distractor should not have increased the target–distractor similarity for the unimodal target group. Note that the tactile vibration was not linked to either of the two response alternatives that the participants had to choose from, and therefore could only affect the resulting compatibility effect by either increasing or decreasing the similarity between the target and the distractor. Across all conditions, the participants showed large compatibility effects; they responded more rapidly and made fewer errors when the target and distractor were mapped onto the same response. The fact that the distractors were highly salient and contained task-relevant information might explain these large compatibility effects. Only the temporal order of the stimuli enabled the selection of the target from the distractor (i.e., participants had to ignore the first stimulus and respond to the second stimulus). The nature of the large compatibility effects will be discussed in more detail in the [General discussion](#). The main finding to emerge from the analysis of the results of Experiment 1a, however, was the difference in the magnitudes of the compatibility effects between bimodal and unimodal distractors. A difference in the sizes of the compatibility effects was observed only in the bimodal target condition. By contrast, in the unimodal target condition, no such difference was observed. Thus, target–distractor similarity was manipulated by response-irrelevant tactile stimulation. According to the contingent capture hypothesis, the participants in the different target conditions applied different top-down sets

(unimodal vs. bimodal). As a consequence, bimodal distractors captured the participants' attention more efficiently than did unimodal distractors because of the increased similarity between the target and the distractor.

As an additional experimental manipulation, the mode of stimulus presentation was manipulated (fixed vs. variable) in Experiment 1a. In the fixed mode of stimulus presentation, the presence or absence of tactile stimulation during the presentation of the visual distractor was entirely predictable for participants. In those blocks with a variable mode of stimulus presentation, the participants were not able to foresee whether the distractor in the subsequent trial might be unimodal or bimodal. More pronounced attentional capture for bimodal than for unimodal distractors was only observed with the variable mode of presentation. Given this result, only an unpredictable tactile stimulation during the presentation of the visual distractor influenced the processing of the current stimulus. By contrast, when the distractor modality (unimodal vs. bimodal) could be foreseen by the participants, the distractor processing was not affected by the absence or presence of a vibrotactile distractor stimulation.

Experiment 1b was designed to replicate the main finding of Experiment 1a—specifically, the difference in the sizes of the compatibility effects for two distractor types (bimodal vs. unimodal). By eliminating the fixed mode of stimulus presentation, the experimental design could be simplified somewhat, and hence all of the independent variables (IVs) in Experiment 1b were manipulated on a within-subjects basis. Consequently, any decrease of variance due to the elimination of the between-subjects manipulation should result in increased statistical power. Thus, it was assumed that, in addition to the highly significant one-sample  $t$  test, the critical three-way interaction should now reach two-tailed significance, whereas the comparable interaction of Experiment 1a (the four-way interaction) reached only one-tailed significance.

## Experiment 1b

### Method

**Participants** Twenty students (17 women, three men; mean age 22 years) served as the participants. All of them reported normal or corrected-to-normal vision and no impairments of somatosensory perception.

**Design** The design of Experiment 1b changed slightly from that of Experiment 1a: Target modality was now manipulated on a within-subjects basis. To achieve this, the fixed mode of stimulus presentation was removed from the design, and only the variable mode of presentation was used in Experiment 1b. Thus, the participants were tested in a  $2 \times 2 \times 2$  factorial design with Response Compatibility (compatible vs.

incompatible), Distractor Modality (unimodal vs. bimodal), and Target Modality (unimodal vs. bimodal) as factors.

**Apparatus and materials** These were identical to those used in Experiment 1a.

**Procedure** The sequence of events for each trial was exactly the same as in Experiment 1a. Each participant had to work through two consecutive blocks of experimental trials, and target modality (bimodal vs. unimodal) was manipulated between the two experimental blocks. In the unimodal target block, the visual targets were never accompanied by vibrotactile stimulation, whereas in the bimodal target block, the visual targets were always accompanied by a vibrotactile stimulus. The sequence of experimental blocks was counterbalanced across participants. Each of the two blocks of trials was initiated by 24 practice trials, in order to allow the participants to adapt to the new target properties. Both experimental blocks comprised 336 experimental trials, so the number of trials per condition was identical to that in the previous experiment.

*Results*

The same criteria as in the previous experiment were used for data trimming. Due to these restrictions, 8.1 % of all trials were excluded from the RT analyses. The mean RTs and error rates are depicted in Table 2.

**RTs** The corrected RTs were submitted to a 2 (Response Compatibility: compatible vs. incompatible) × 2 (Distractor Modality: bimodal vs. unimodal) × 2 (Target Modality: bimodal vs. unimodal) MANOVA with Pillai’s trace as the criterion. Consistent with the findings of Experiment 1a, the main effects of response compatibility,  $F(1, 19) = 124.488, p < .001, \eta_p^2 = .868$ , and of distractor modality,  $F(1, 19) = 22.78, p < .001, \eta_p^2 = .545$ , were significant: The participants responded more rapidly when the distractor and the target were both linked to the same response, and they also responded more rapidly after a bimodal than after a unimodal distractor. More importantly, the three-way interaction

reached significance,  $F(1, 19) = 5.379, p = .032, \eta_p^2 = .221$ . This interaction was further analyzed by means of two one-sample *t* tests. In line with the results of Experiment 1a, the sizes of the compatibility effects only differed in the bimodal target condition,  $M_{Diff} = 16 \text{ ms} (SD = 24 \text{ ms}), t(19) = 2.985, p = .008$  (bimodal distractors,  $M = 136 \text{ ms}, SD = 53 \text{ ms}$ , vs. unimodal distractors,  $M = 120 \text{ ms}, SD = 53 \text{ ms}$ ). No such difference in the sizes of the compatibility effects was observed in the unimodal target condition,  $M_{Diff} = 3 \text{ ms} (SD = 19 \text{ ms}), t(19) = 0.677, p = .506$  (bimodal distractors,  $M = 128 \text{ ms}, SD = 53 \text{ ms}$ , vs. unimodal distractors,  $M = 124 \text{ ms}, SD = 54 \text{ ms}$ ) (see Fig. 4).

**Error rates** The same MANOVA with the error rates as the dependent variable revealed a significant main effect of response compatibility,  $F(1, 19) = 19.613, p < .001, \eta_p^2 = .508$ : The participants made fewer errors in those trials in which the targets and distractors were both mapped to the same response. Additionally, the Distractor Modality × Target Modality interaction was also significant,  $F(1, 19) = 5.760, p = .027, \eta_p^2 = .233$ . In the bimodal target block, the participants made more errors after the presentation of a bimodal distractor ( $M = 4.6 \%$ ) than after the presentation of a unimodal distractor (3.9 %). By contrast, the participants made fewer errors after the presentation of a bimodal distractor (4.1 %) than after the presentation of a unimodal distractor (5.0 %) in the unimodal distractor condition. However, this interaction was not further specified by the three-way interaction,  $F < 1$ .

*Discussion*

Experiment 1b was, in most respects, an exact replication of Experiment 1a. The aim was to provide additional support for the claim of contingent crossmodal capture on a within-subjects basis. The results of Experiment 1b fully confirmed the main finding from Experiment 1a—namely that in the bimodal target condition, larger compatibility effects were observed after the presentation of a bimodal distractor than after the presentation of a unimodal distractor. Once again, no such difference in the sizes of the compatibility effects was

**Table 2** Mean RTs (in milliseconds) and mean error rates (as percentages, in parentheses) as a function of response compatibility (incompatible vs. compatible), target modality (bimodal vs. unimodal), and distractor modality (bimodal vs. unimodal) in Experiment 1b

	Bimodal Target		Unimodal Target	
	Bimodal Distractor	Unimodal Distractor	Bimodal Distractor	Unimodal Distractor
Incompatible	534 (7.3)	544 (6.5)	537 (1.5)	543 (2.7)
Compatible	398 (2.0)	424 (1.3)	409 (6.6)	419 (7.3)
Compatibility effect	136 (5.3)	120 (5.2)	128 (5.1)	124 (4.6)

Compatibility effects were computed as the difference between response-incompatible and response-compatible trials.



observed within the unimodal target condition; the presentation of an additional tactile stimulus during the presentation of the distractor did not affect the size of the compatibility effect when the participants were searching for a unimodal target.

An interesting aspect of Experiment 1a was the lack of a boost in the magnitude of the compatibility effect for the bimodal distractors in the fixed mode of stimulus presentation. In their recent study, Mast and Frings (2014) showed that participants adopt efficient top-down sets. Thus, only those features that provide additional information for efficient task performance are implemented into the top-down sets. That could be those features that define the correct response (*response features*), but also could be those features that help to select the targets from among the distractors (*selection features*). For the variable mode of stimulus presentation in Experiment 1a, the tactile stimulus provided helpful information to separate the target from the distractor, at least on half of the trials. By contrast, when the target and the distractor were always presented together with a tactile stimulus, the tactile stimulus did not provide any information that might help to separate the target from the distractor. Thus, in the bimodal target condition, when the participants were always confronted with only unimodal distractors (i.e., the blockwise presentation mode in Exp. 1a with bimodal targets), the tactile stimulus signaling the presence of the target should have been implemented into the top-down control sets—that is, because the tactile feature helped to separate the target from the distractor. The unimodal distractors, however, never matched that tactile feature of the top-down control sets, and consequently could not boost the resulting compatibility effect (see Pratt & McAuliffe, 2002).

The observed differences in the sizes of the compatibility effects in Experiment 1 are assumed to reflect differences in the potentials of unimodal and bimodal distractors to involuntarily capture the participant's attention. However, one might argue that contingent crossmodal capture is not the only explanation that might account for the results obtained here. In the supraliminal response-priming task utilized here, the targets and the distractors shared all basic physical properties (i.e., shape, color, size, and location); thus, only the temporal sequence of stimulus presentation allowed the participants to tell the target from the distractor. Consequently, the participants had to process the distractors to a certain level in order to identify the subsequent stimulus as the target. By contrast, in the exogenous cuing task (e.g., Folk et al., 1992)—the paradigm that is typically used to investigate contingent capture—the cues (e.g., four dots that encircle one of the possible target locations) and the targets (e.g., a “T” or an “=” symbol) differ regarding to their basic physical properties (e.g., shape); typically, an absolute feature separates the targets and the distractors (e.g., shape). To address this thorny issue, an absolute visual *selection feature* was implemented in Experiment 2: The targets and the distractors were presented either in the

same shape (shape congruent) or in different shapes (shape incongruent) throughout an entire block of trials. In the shape-incongruent condition, the targets and distractors always differed according to their basic physical properties. Thus, the processing of the distractor was no longer necessary when it came to identifying the target in the shape-incongruent condition. If the sizes of the compatibility effects still varied for the bimodal and unimodal distractors, that finding would underpin the claim of involuntary contingent crossmodal capture. In addition, shape-congruent trials were utilized as a control condition.

## Experiment 2

### Methods

**Participants** Twenty students (15 women, five men; mean age 22 years) were tested. All of the participants reported normal or corrected-to-normal vision and no impairments of somatosensory perception.

**Design** The participants were tested in a  $2 \times 2 \times 2$  design with Response Compatibility (compatible vs. incompatible), Distractor Modality (unimodal vs. bimodal), and Shape Congruency (congruent vs. incongruent) as factors.

**Apparatus and materials** These were in most respects identical to the same aspects of Experiment 1a and 1b. Yet, one major change was made with respect to the visual stimuli: The targets were still circles ( $1.72^\circ$ ); however, the distractors could be either the same shape or else squares ( $1.72^\circ$  side length).

**Procedure** The trial sequence was identical to that used in Experiment 1. Two experimental blocks were applied that differed only in a manipulation of the shape of the distractor stimuli. The distractor shapes were either congruent or incongruent for an entire block. The sequence of blocks was counterbalanced across participants. Each experimental block comprised 216 trials, of which half were response compatible and half, response incompatible. The distractor modality (unimodal vs. bimodal) was manipulated orthogonally. The experimental blocks were preceded by 48 practice trials each.

### Results

The same rules for data trimming were applied as in the previous experiment, leading to 8.2 % of all trials being excluded from the further analyses. The mean RTs and error rates are depicted in Table 3.

**RTs** A 2 (Response Compatibility: compatible vs. incompatible)  $\times$  2 (Distractor Modality: unimodal vs. bimodal)  $\times$  2 (Shape Congruency: congruent vs. incongruent) MANOVA

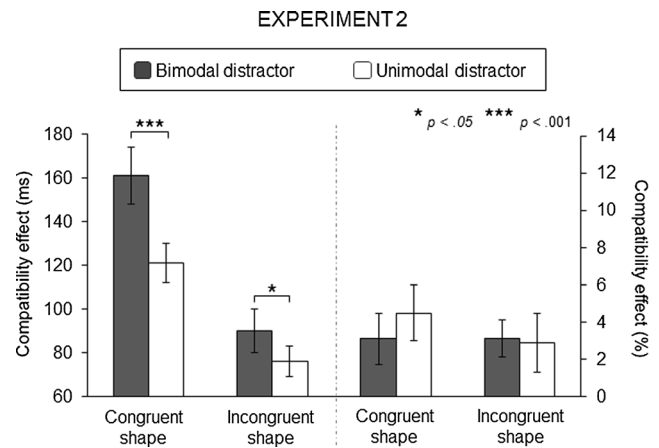
**Table 3** Mean RTs (in milliseconds) and mean error rates (as percentages, in parentheses) as a function of response compatibility (incompatible vs. compatible), distractor modality (bimodal vs. unimodal), and shape congruency (congruent vs. incongruent) in Experiment 2

Shape Congruency		Distractor Modality	
		Bimodal	Unimodal
Congruent	Incompatible	586 (5.7)	582 (5.9)
	Compatible	426 (2.7)	461 (2.9)
	Comp. effect	161 (3.1)	121 (3.1)
Incongruent	Incompatible	531 (6.0)	539 (6.2)
	Compatible	442 (1.5)	463 (3.3)
	Comp. effect	90 (4.5)	76 (2.9)

Only bimodal targets were utilized in Experiment 2. Compatibility effects (Comp. effect) were computed as the difference between response-incompatible and response-compatible trials.

was conducted with Pillai’s trace as the criterion and correct RTs as the dependent variable. The key analyses were the interactions between response compatibility and the two remaining factors—namely, the modulations of the compatibility effects. Both the Response Compatibility × Distractor Modality interaction,  $F(1, 19) = 21.907, p < .001, \eta_p^2 = .536$ , and the Response Compatibility × Shape Congruency interaction,  $F(1, 19) = 59.808, p < .001, \eta_p^2 = .759$ , reached statistical significance. The sizes of the compatibility effect varied as a function of the distractor shape and distractor modality. What is more, the three-way interaction was significant, too,  $F(1, 19) = 7.453, p = .013, \eta_p^2 = .282$ . That is, shape-congruent distractors led to larger compatibility effects than did shape-incongruent distractors. Intriguingly, the sizes of the compatibility effects caused by congruent and incongruent distractors differed as a function of distractor modality (see Fig. 5). For the shape-congruent condition, larger compatibility effects were documented after the presentation of a bimodal distractor ( $M = 161$  ms,  $SD = 56$  ms) than after the presentation of a unimodal distractor ( $M = 121$  ms,  $SD = 45$  ms),  $M_{Diff} = 40$  ms ( $SD = 39$  ms),  $t(19) = 4.504, p < .001$ . A similar pattern of results was observed for the shape-incongruent distractors, as well. That is, larger compatibility effects were elicited by bimodal distractors ( $M = 90$  ms,  $SD = 38$  ms) than by unimodal distractors ( $M = 76$  ms,  $SD = 34$  ms),  $M_{Diff} = 14$  ms ( $SD = 25$  ms),  $t(19) = 2.435, p = .025$ .

We observed a significant main effect of distractor modality,  $F(1, 19) = 33.626, p < .001, \eta_p^2 = .639$ : Participants responded more rapidly after the presentation of a bimodal distractor than after a unimodal distractor. Finally, the MANOVA revealed a significant main effect of shape congruency,  $F(1, 19) = 5.464, p = .031, \eta_p^2 = .223$ , with participants responding more slowly after the presentation of a shape-congruent distractor than after



**Fig. 5** RT (in milliseconds; on the left side) and error rate (as percentages; on the right side) compatibility effects in Experiments 2 as a function of distractor modality and shape congruency. The error bars depict the standard errors of the means.

trials with a shape-incongruent distractor. The Distractor Modality × Shape Congruency interaction did not reach significance,  $F(1, 19) < 1$ .

**Error rates** The same MANOVA as for RTs was used to analyze the error rates. Only the main effect of response compatibility reached significance,  $F(1, 19) = 9.088, p = .007, \eta_p^2 = .324$ . The participants responded more accurately when the target and distractor were mapped onto the same response than when they were mapped onto different responses.

*Discussion*

The second experiment was designed to investigate whether the boost in the size of the compatibility effects observed for bimodal distractors, as compared to unimodal distractors, results from involuntary distractor processing or reflects strategic stimulus processing due to the task relevance of the distractors. That is, in Experiment 1, the target and the distractor shared all of their basic physical properties (i.e., same shape, same location, and same size). Hence, only the temporal sequence of stimulus presentation enabled the participants to select the target (the second stimulus in each trial) instead of the distractor (the first stimulus in each trial). Thus, the distractors needed to be processed in order to identify the target. To address this issue, an absolute visual selection feature (shape) was implemented in Experiment 2. In the shape-incongruent condition, the targets and the distractors had different shapes throughout the entire block. Consequently, the participants only had to respond to the “circles” and ignore the “squares.”

The data from Experiment 2 indicated that participants benefited from the implementation of multiple selection features. That is, the largest compatibility effects were observed following the presentation of bimodal shape-congruent

distractors—that is, distractors that were very similar to the target. By contrast, the smallest compatibility effects were observed following the presentation of unimodal shape-incongruent distractors—that is, distractors that were very dissimilar from the target. The results of Experiment 2 are in line with the assumption that a decline in target–distractor similarity should decrease the potential of a distractor to attract attention automatically (see Ansorge & Heumann, 2003; Ansorge & Heumann, 2004; Mast & Frings, 2014).

Most important, however, is the observation that even in the shape-incongruent condition, larger compatibility effects were found after the presentation of a bimodal distractor than after the presentation of a unimodal distractor. Note that the participants were clearly able to separate the target from the distractor due to the shape of the stimulus. Still, the results indicated that bimodal distractors received more processing than did the unimodal distractors. Therefore, the present results nicely match with the predictions of contingent crossmodal capture hypothesis. That is, the bimodal shape-congruent distractors were most efficient in capturing the participants' attention and received more processing resources because of their high similarity to the target. Weaker attentional capture was observed following target-dissimilar distractors, and the lowest for unimodal shape-incongruent distractors.

In contrast to the results of Experiment 1a, here the participants made use of shape congruency even with a fixed mode of stimulus presentation. The lack of boosting in compatibility effects that was observed for bimodal distractors with the fixed mode of stimulus presentation in Experiment 1a might have been attributable to the fact that the tactile stimulation did not provide any helpful information (i.e., the distractors and the targets were both accompanied by tactile stimulation). Thus, the participants might have suppressed information from the tactile modality for the block (see Mast et al., 2014, for an imbalance between tactile and visual distractor processing). By contrast, information presented visually (the distractors) was much harder to suppress because it was presented in the target modality. Another fundamental difference between the manipulation of distractor modality and the manipulation of shape congruency is the fact that a change in a visual feature directly affects the visual target, whereas the distractor modality was manipulated by means of the on- and offset of an additional, tactile stimulus. Further research will therefore be needed in order to investigate whether the pattern of results could be changed if the primary task were tactile and similarity were manipulated by an additional visual stimulus. Note, however, that such an interpretation does not challenge the contingent crossmodal capture interpretation of the present results. Instead, it triggers the question of when and how the information from a modality other than the target modality is integrated into top-down sets.

Although the main finding of Experiment 1 was replicated, still another issue needs to be considered. In all of the experiments reported so far, the participants' responses to the targets were more rapid after the presentation of a bimodal distractor than after the presentation of a unimodal distractor. The differences in mean RTs might indicate that the presentation of an additional tactile stimulation during distractor presentation might have amplified the distractor signal, and therefore operated as a “wake-up” call for the participants (see Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008, 2009). Consequently, the participant might have paid more attention to the current input (the distractor), which may have resulted in enhanced distractor processing. Yet an alerting-based explanation for the data pattern observed in Experiment 1 still depends on top-down control. That is, more rapid responses after the presentation of a bimodal distractor were observed for both the unimodal and bimodal target conditions, but only in the bimodal target condition were responses accelerated and differences in the sizes of the compatibility effects observed.<sup>3</sup>

To address the differences in the alerting potentials that were observed for bimodal and unimodal distractors in Experiments 1 and 2, a third experiment was conducted, with different tactile patterns being presented instead of the mere presence versus absence of a tactile stimulus. Thus, in Experiment 3, the presentation of the visual target was always accompanied by a specific tactile pattern. In half of the trials, the visual distractors and the targets were accompanied by the same tactile pattern (congruent trials). In the remaining trials, the tactile patterns differed for the presentation of the visual distractor and presentation of the visual target (incongruent trials). The utilization of different tactile patterns enabled a comparison of bimodal distractors that were either congruent (target similar) or incongruent (target dissimilar) to the tactile target pattern. However, both modalities were always stimulated, and therefore the alerting effects of a bimodal distractor should have been equalized. Once again, larger compatibility effects were expected to be found for those trials with a congruent tactile stimulation during visual distractor presentation than for those in which distractors were accompanied by an incongruent tactile stimulation. The different tactile patterns were designed by means of a manipulation of the stimulus intensity (see Mast et al., 2014).

### Experiment 3

#### Method

**Participants** Forty students (31 women, nine men; mean age 22 years) served as the participants in Experiment 3. All of them reported normal or corrected-to-normal vision and no impairments of somatosensory perception.

<sup>3</sup> We thank an anonymous reviewer for suggesting this alternative explanation.

**Design** The participants were tested in a  $2 \times 2$  design with Response Compatibility (compatible vs. incompatible) and Tactile Congruency (congruent vs. incongruent) as within-subjects factors.

**Apparatus and materials** The technical equipment was the same as in the previous experiments. However, in order to use different tactile patterns, the tactile stimulus set had to be adapted. Two tactile patterns were designed by using tactile stimuli that differed in their intensity (weak vs. strong). The presentation time for the distractors and the targets was increased to 67 ms, and the intensity of the tactile target stimulation was kept constant, either strong or weak, for each participant. By contrast, the intensity of the tactile stimulation during distractor presentation was randomly manipulated. Note that the tactile target patterns were balanced across participants.

**Procedure** The sequence of events for each trial was identical to that in the previous experiments, aside from the slight increase in the presentation times for targets and distractors. Before starting the actual experiment, the participants had to work their way through 64 training trials. Only one experimental block was presented in Experiment 3. This block comprised 336 trials (84 compatible with congruent tactile stimulation, 84 incompatible with congruent tactile stimulation, 84 compatible with incongruent tactile stimulation, and 84 incompatible with incongruent tactile stimulation).

### Results

Data trimming followed the same rules as in all of the experiments above. Due to these restrictions, 8.0 % of all trials were excluded from the RT analyses. The data of one participant had to be excluded from the analyses due to his extremely high error rate (19.64 %, as compared to the sample's mean error rate of 4.0 %) and his extremely slow responses (801 ms, as compared to the sample's mean RT of 495 ms). The mean RTs and error rates are shown in Table 4.

**RTs** The RTs from Experiment 3 were submitted to a 2 (Response Compatibility: compatible vs. incompatible)  $\times$  2 (Tactile Congruency: congruent vs. incongruent) MANOVA with Pillai's trace as the criterion. A main effect of response compatibility was observed,  $F(1, 37) = 280.84, p < .001, \eta_p^2 = .881$ : The participants responded more rapidly when the target and the distractor were mapped onto the same response than when they were mapped onto different responses. As expected, the main effect of tactile congruency did not reach significance,  $F(1, 37) < 1$ . That is, the participants' mean RTs did not differ as a function of tactile congruency. However, the Response Compatibility  $\times$  Tactile Congruency interaction was significant,  $F(1, 38) = 5.362, p = .026, \eta_p^2 = .124$ , in that the sizes of the compatibility effects differed as a function of whether the

**Table 4** Mean RTs (in milliseconds) and mean error rates (as percentages, in parentheses) as a function of response compatibility (incompatible vs. compatible) and tactile congruency (congruent vs. incongruent) in Experiment 3

	Tactile Congruency	
	Congruent Distractor	Incongruent Distractor
Incompatible	557 (7.7)	552 (7.1)
Compatible	430 (3.6)	439 (2.4)
Compatibility effect	126 (4.1)	114 (4.7)

Compatibility effects were computed as the difference between response-incompatible and response-compatible trials.

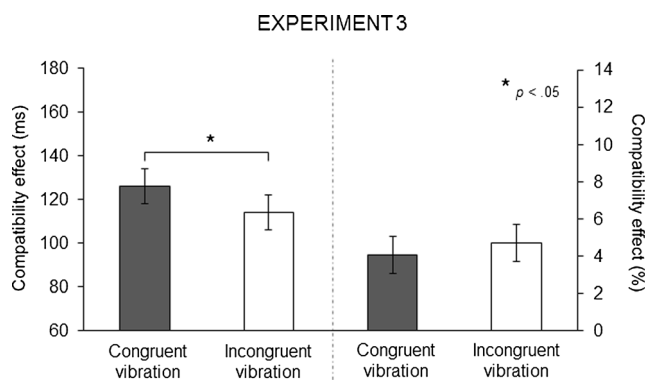
distractor was presented together with a congruent or an incongruent tactile stimulation (see Fig. 6). As predicted, larger compatibility effects were observed in the congruent tactile condition ( $M = 126$  ms,  $SD = 48$  ms) than in the incongruent tactile condition ( $M = 114$  ms,  $SD = 48$  ms).<sup>4</sup>

**Error rates** The error rates were submitted to the same  $2 \times 2$  MANOVA. Only the main effect of response compatibility reached significance,  $F(1, 38) = 21.781, p < .001, \eta_p^2 = .371$ ; the participants made fewer errors when the target and the distractor were both mapped onto the same response than when they were mapped onto different responses. For the error rates, neither the main effect of tactile congruency,  $F(1, 38) = 3.025, p = .09, \eta_p^2 = .074$ , nor the Response Compatibility  $\times$  Tactile Congruency interaction,  $F(1, 38) < 1$ , reached significance. The results revealed that neither the mean error rates nor the resulting compatibility effects were affected by tactile stimulation.<sup>5</sup>

<sup>4</sup> The same MANOVA was conducted with Tactile Stimulus Intensity as the between-subjects factor, to test whether the difference in the sizes of the compatibility effects was systematically affected by the intensity of the tactile stimulation (strong vs. weak). Neither the main effect of tactile stimulus intensity,  $F(1, 37) < 1$ , nor any of the interactions with tactile stimulus intensity reached significance [Tactile Congruency  $\times$  Tactile Stimulus Intensity,  $F(1, 37) < 1$ ; Response Compatibility  $\times$  Tactile Stimulus Intensity,  $F(1, 37) < 1$ ; Response Compatibility  $\times$  Tactile Congruency  $\times$  Tactile Stimulus Intensity,  $F(1, 37) = 1.226, p = .275, \eta_p^2 = .032$ ].

<sup>5</sup> The same MANOVA tested with Tactile Stimulus Intensity as the between-subjects factor revealed no significant results—neither a main effect of tactile stimulus intensity,  $F(1, 37) = 2.357, p = .133, \eta_p^2 = .06$ , nor any interactions [Tactile Congruency  $\times$  Target Intensity,  $F(1, 37) = 2.724, p = .107, \eta_p^2 = .069$ ; Response Compatibility  $\times$  Target Intensity,  $F(1, 37) = 1.626, p = .210, \eta_p^2 = .042$ ; Response Compatibility  $\times$  Tactile Congruency  $\times$  Target Intensity,  $F(1, 37) < 1$ ].





**Fig. 6** RT (in milliseconds; on the left side) and error rate (as percentages; on the right side) compatibility effects in Experiments 3 as a function of tactile congruency (in this experiment, only bimodal targets were utilized). The error bars depict the standard errors of the means.

### Discussion

Experiment 3 was conducted to investigate whether the difference in the sizes of the compatibility effects was primarily driven by the alerting potential of a bimodal distractor signal. To test this alternative account, only bimodal distractors were utilized. However, the bimodal distractors differed in that they were presented with either a target-congruent or a target-incongruent tactile stimulus. In line with the previous experimental results, larger compatibility effects were found for those trials in which the target and the distractor were presented with the same tactile stimulus pattern than for trials in which the distractor was presented together with an incongruent tactile stimulus. Thus, the participants were able to set up a specific tactile *selection feature* (i.e., either for a “strong” or a “weak” vibrotactile pattern) as part of the top-down set. According to the specified top-down sets, the potential of a distractor to attract attention varied as a function of tactile congruency; larger compatibility effects were observed when the distractors were accompanied by the same vibration as the targets (target similar) than when the distractors were accompanied by a vibration that differed from the stimulation during target presentation (target dissimilar). Note that whereas the sizes of the compatibility effects differed for trials with congruent and incongruent tactile stimulation, no differences were observed in participants’ mean RTs. Thus, the increased alerting potential caused by bimodal distractors in Experiments 1 and 2 cannot entirely account for the differences in the sizes of the compatibility effects. The results were also not confounded by the fact that participants differed in attending to either a strong or a weak tactile target pattern. Therefore, the results of Experiment 3 provide further support for the existence of contingent crossmodal capture and counter the argument that the differences in the sizes of the compatibility effects in our earlier experiments were driven primarily by an alerting signal due to the bimodal distractor presentation.

### General discussion

To date, research concerning the interaction between top-down and bottom-up control of attention, as embodied by the *contingent capture hypothesis* (Folk et al., 1992), has primarily been conducted within the visual modality. By contrast, research on crossmodal (spatial) attentional control has focused primarily on bottom-up mechanisms, instead (Spence, 2013). The present study was therefore conducted in order to address this gap in the literature and to investigate whether top-down sets could also be applied in a multisensory task environment (involving visual and tactile stimulation). During each trial, the target and distractor (both visual) were mapped either onto the same response (compatible trials) or onto different responses (incompatible trials) and were presented sequentially in the same spatial location. In order to test whether top-down sets could contain features from different sensory modalities, target–distractor similarity was manipulated by means of a response-irrelevant vibrotactile stimulus during distractor and/or target presentation. In the bimodal target condition, the visual target was always presented together with a vibrotactile stimulus, whereas in the unimodal target condition, no such vibrotactile stimulation was presented during the presentation of the target.

In all of the experiments reported here, the participants responded more rapidly, and with fewer errors, in those trials in which the target and distractor were both linked to the same response than in those in which target and distractor were associated with different responses; in other words, a compatibility effect was observed. Although throughout the course of the experiments the distractor did not predict the subsequent target stimulus, the distractors still had a strong impact on participants’ responses to the subsequently presented target. In all conditions, large main effects of compatibility were observed (greater than 100 ms across all conditions). The magnitude of these effects is not surprising, given the fact that each distractor (independent of the tactile stimulation) was highly salient (a sudden onset stimulus, exclusively presented on a black background; see Wolfe, 1994, for the guided search theory) and shared most of its perceptual attributes with the target stimulus—namely shape, location, abrupt onset, and most important, the response-relevant feature (green vs. red). In fact, in order to identify the target stimulus, it was necessary to recognize the distractor stimulus (i.e., as the first stimulus), because only the relative temporal order (responding to the second of the two stimuli) allowed the participants to discriminate between the target and the distractor (in Exps. 1 and 3). Hence, one might argue that the distractors did not actually capture attention, but instead were actively processed by the participants in order to identify the target. If strategic distractor processing were the origin of the differences in the sizes of the compatibility effects that are the main finding of the present study, that would eliminate contingent crossmodal capture as

an appropriate theoretical framework. The results of Experiment 2, however, indicate that the findings reported here do not exclusively reflect voluntary task strategies, but rather involuntary distractor processing. In the shape-incongruent condition, the targets and the distractors could always be separated due to their basic visual properties (i.e., the stimulus shape). Still, large compatibility effects were observed even for shape-incongruent distractors.

However, the most important findings to emerge from the present study were that the size of the compatibility effect varied as a function of the similarity between the target and the distractor and that target–distractor similarity can also be manipulated by stimulation from different sensory modalities. In the bimodal target condition of Experiments 1a and b, the appearance of the target was always combined with the presentation of a tactile stimulus. We assumed that the continual co-occurrence of the visual target with the tactile stimulus would have led to an integration of the vibrotactile feature into the (otherwise purely visual) top-down set. In line with the assumption of extended multisensory top-down sets (cf. Folk et al., 1992), bimodal distractors led to enhanced capture of participants' attention, indicated by the increased compatibility effects relative to unimodal distractors (see Ansorge & Heumann, 2003, 2004, for similar effects in a unisensory spatial-cuing task, but with color as the feature used to vary target–distractor similarity). In the unimodal target condition, the presence of the target was never accompanied by a tactile stimulus, and hence the top-down sets should only have been composed of visual features. In line with the reasoning outlined here, the size of the compatibility effects in the unimodal target conditions of Experiments 1a and 1b was not affected by the modality of the cue.

One might be tempted to explain the present results by means of different bottom-up processes instead of through the assumption of any mediating top-down sets. In their review, Talsma, Senkowski, Soto-Faraco, and Woldorff (2010) assumed that multisensory events are more salient than unisensory events (cf. Spence, 2010); consequently, multisensory events tend to be more efficient in capturing attention and to receive more processing resources than unisensory events (cf. Van der Burg et al., 2008, 2009; see also Ngo & Spence, 2010; see Spence, 2010, for a review of the literature on crossmodal selective attention). The results of the bimodal distractor conditions in Experiments 1 and 2 could, at first glance, be explained by differences in distractor saliency instead of differences in target–distractor similarity. Thus, bimodal distractors may have been experienced as more salient than the opposing unimodal distractors, and therefore have captured attention more efficiently. That interpretation, however, contrasts with the results of the unimodal target condition, in which a difference in the sizes of the compatibility effects caused by bimodal distractors was observed only when the targets were associated with the vibrotactile stimulus, but

not in the unimodal target condition. In the two target modality conditions (bimodal vs. unimodal target), perceptually identical distractors (bimodal vs. unimodal) were utilized. According to the salience-based explanation of the results outlined here, the bimodal distractors should have captured the participants' attention more efficiently than unimodal distractors, irrespective of the target modality. More importantly, the saliency-based explanation was tested in Experiment 3 with distractors that were always bimodal (so that any alerting or salience explanation simply could not be applied). Therefore, we argue that the results of the present study cannot be explained in terms of saliency differences for bimodal and unimodal distractors, but must instead reflect the interplay of multisensory top-down sets and bottom-up mechanisms.

The contingent crossmodal capture hypothesis presented here provides new implications for the original contingent capture account postulated by Folk et al. (1992), but also for its revised version, the *displaywide contingent capture hypothesis* of Gibson and Kelsey (1998; see Burnham, 2007, for a review). The two accounts are strongly related, and both underline the importance of attentional control settings for the occurrence of attentional capture. Yet the two accounts differ in which features they presume to be integrated into the participants' top-down sets. On the one hand, the original contingent capture account predicts that those features that are crucial for the identification and localization of the target stimulus will be implemented into the participants' top-down sets. Gibson and Kelsey, however, emphasized that, in parallel with the participants' primary task (i.e., the localization and identification of the target stimulus), participants monitor the visual input for the onset of the target display. According to this view, participants implement those visual features that help to identify the target stimulus, but they also implement those features that signal the onset of the target display. Note that the results of the present study are consistent with both accounts. In line with the classic contingent capture hypothesis, the presentation of the tactile stimulation might have been implemented in the participants' top-down sets because it was attributed as a tactile feature of a multisensory target stimulus (i.e., an integration of the visual and tactile information into a multisensory object representation; see Iordanescu, Grabowecky, Franconeri, Theeuwes, & Suzuki, 2010; Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008). By contrast, the displaywide contingent capture hypothesis would explain the results by arguing that the tactile stimulus was integrated into the participants' top-down sets because its appearance indicated the onset of the target display. Even though the present study might not differentiate between a *stimulus-centered* and a *display-centered* contingent capture account, the utilization of crossmodal experimental settings will provide new arguments for either of the two accounts in future projects.

Taken together, the results reported here are the first to provide clear evidence that the concept of top-down sets, as postulated by the contingent capture hypothesis, can be extended to a multisensory task environment, or at least to a visuotactile task environment. The compilation of top-down sets is based on the associations between different sensory modalities during target presentation. Top-down sets further have a crucial impact on bottom-up processing. The overlap between the features of a distractor and the features of the current top-down set determines the distractor's potential to involuntarily capture participants' attention. More efficient attentional capture (i.e., larger compatibility effects) was observed for bimodal than for unimodal distractors, but only when the tactile information was correlated with the appearance of the target stimulus.

**Author note** The research reported in this article was supported by a grant of the Deutsche Forschungsgemeinschaft to C.F. and C.S. (FR 2133/5-1). We thank Caroline Bermes for drawing Fig. 3, as well as Nina Schmitz and Simon Merz for their help with data collection.

## 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. doi:10.3758/APP.72.2.342
- 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. doi:10.1037/0096-1523.29.5.937
- Ansorge, U., & Heumann, M. (2004). Peripheral cuing by abrupt-onset cues: The influence of color in S–R corresponding conditions. *Acta Psychologica*, *116*, 115–143. doi:10.1016/j.actpsy.2004.01.001
- Ansorge, U., Heumann, M., & Scharlau, I. (2002). Influences of visibility, intentions, and probability in a peripheral cuing task. *Consciousness and Cognition*, *11*, 528–545. doi:10.1016/S1053-8100(02)00026-0
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*, 437–443. doi:10.1016/j.tics.2012.06.010
- Burnham, B. R. (2007). Displaywide visual features associated with a search display's appearance can mediate attentional capture. *Psychonomic Bulletin & Review*, *14*, 392–422. doi:10.3758/BF03194082
- Eimer, M., Kiss, M., Press, C., & Sauter, D. (2009). The roles of feature-specific task set and bottom-up salience in attentional capture: An ERP study. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1316–1328. doi:10.1037/a0015872
- Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics*, *64*, 741–753. doi:10.3758/BF03194741
- Folk, C. L., Leber, A. B., & Egeth, H. E. (2008). Top-down control settings and the attentional blink: Evidence for nonspatial contingent capture. *Visual Cognition*, *16*, 616–642. doi:10.1080/13506280601134018
- Folk, C. L., & Remington, R. W. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 847–858. doi:10.1037/0096-1523.24.3.847
- 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. doi:10.1037/0096-1523.18.4.1030
- Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 317–329. doi:10.1037/0096-1523.20.2.317
- Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 699–706. doi:10.1037/0096-1523.24.3.699
- Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception, & Psychophysics*, *72*, 1736–1741. doi:10.3758/APP.72.7.1736
- Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, *15*, 548–554. doi:10.3758/PBR.15.3.548
- Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long & A. Baddeley (Eds.), *Attention and performance* (pp. 187–203). Hillsdale: Erlbaum.
- Kiesel, A., Kunde, W., & Hoffmann, J. (2007). Mechanisms of subliminal response priming. *Advances in Cognitive Psychology*, *3*, 307–315. doi:10.2478/v10053-008-0032-1
- Mast, F., & Frings, C. (2014). The impact of the irrelevant: The task environment modulates the impact of irrelevant features in response selection. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 2198–2213. doi:10.1037/a0038182
- Mast, F., Frings, C., & Spence, C. (2014). Response interference in touch, vision, and crossmodally: Beyond the spatial dimension. *Experimental Brain Research*, *232*, 2325–2336. doi:10.1007/s00221-014-3929-6
- Neumann, O., & Klotz, W. (1994). Motor responses to nonreportable, masked stimuli: In which is the limit of direct parameter specification? In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 123–150). Cambridge: MIT Press.
- Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, *72*, 1654–1665. doi:10.3758/APP.72.6.1654
- Nobre, A. C. (2001). Orienting attention to instants in time. *Neuropsychologia*, *39*, 1317–1328. doi:10.1016/S0028-3932(01)00120-8
- Nobre, A. C., & Coull, J. T. (Eds.). (2010). *Attention and time*. Oxford: Oxford University Press.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25. doi:10.1080/00335558008248231
- Pratt, J., & McAuliffe, J. (2002). Determining whether attentional control settings are inclusive or exclusive. *Perception & Psychophysics*, *64*, 1361–1370. doi:10.3758/BF03194778
- Rauschenberger, R. (2003). When something old becomes something new: Spatiotemporal object continuity and attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 600–615. doi:10.1037/0096-1523.29.3.600
- Shore, D. I., & Simic, N. (2005). Integration of visual and tactile stimuli: Top-down influences require time. *Experimental Brain Research*, *166*, 509–517. doi:10.1007/s00221-005-2391-x
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*, 182–200. doi:10.1111/j.1749-6632.2010.05440.x

- Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296, 31–49. doi:10.1111/nyas.12121
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, 14, 400–410. doi:10.1016/j.tics.2010.06.008
- Theeuwes, J. (1991). Exogenous and endogenous control of visual attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, 49, 83–90. doi:10.3758/BF03211619
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, 51, 599–606. doi:10.3758/BF03211656
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135, 77–99. doi:10.1016/j.actpsy.2010.02.006
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.
- Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin*, 135, 452–477. doi:10.1037/a0015329
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1053–1065. doi:10.1037/0096-1523.34.5.1053
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile–visual synchrony increases visual saliency. *Neuroscience Letters*, 450, 60–64. doi:10.1016/j.neulet.2008.11.002
- van der Lubbe, R. H. J., & Van der Helden, J. (2006). Failure of the extended contingent attentional capture account in multimodal settings. *Advances in Cognitive Psychology*, 2, 255–267. doi:10.2478/v10053-008-0060-x
- Wesslein, A. K., Spence, C., & Frings, C. (2014). Vision affects tactile target and distractor processing even when space is task-irrelevant. *Frontiers in Psychology*, 5, 84. doi:10.3389/fpsyg.2014.00084
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238. doi:10.3758/BF03200774
- Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 601–621. doi:10.1037/0096-1523.10.5.601