Citation for final published version:

Yamaguchi, Motonori, Valji, Ashvanti and Wolohan, Felicity 2018. Top-down contributions to attention shifting and disengagement: a template model of visual attention. Journal of Experimental Psychology: General 147 (6) , pp. 859-887. 10.1037/xge0000393

Publishers page: http://dx.doi.org/10.1037/xge0000393

Yamaguchi, M., Valji, A., & Wolohan, F. D. A. (in press). Top-down contributions to attention shifting and disengagement: A template model of visual attention. *Journal of Experimental Psychology: General*.

Top-Down Contributions to Attention Shifting and Disengagement:

A Template Model of Visual Attention

Motonori Yamaguchi[1], Ashvanti Valji[2], and Felicity D. A. Wolohan[1]

[1]Edge Hill University, Ormskirk, UK

[2]Cardiff University, Cardiff, UK

Author Notes

Motonori Yamaguchi and Felicity D. A. Wolohan, Department of Psychology, Edge Hill University, Ormskirk, the United Kingdom; Ashvanti Valji, School of Psychology, Cardiff University, the United Kingdom. Correspondence concerning this article should be addressed to Motonori Yamaguchi, Department of Psychology, Edge Hill University, St Helens Road, Ormskirk, Lancashire, L39 4QP, UK. E-mail: yamagucm@edgehill.ac.uk.

The present study was supported in part by the Research Investment Fund from Edge Hill University. We thank Trish Van Zandt for her technical assistance for the Wald diffusion model, Gordon D. Logan for discussions on modeling, Yu-Chin Chiu for suggesting the Bayesian model selection for group data, and David Lilley for help on data collection.

The data reported in the present article have not been disseminated in any other form. The experimental data reported in the present article are available from the project page in the Open Science Framework website (https://osf.io/dzs2k/).

Abstract

Two separate systems are involved in the control of spatial attention; one that is driven by a goal, and the other that is driven by stimuli. While the goal- and stimulus-driven systems follow different general principles, they also interplay with each other. However, the mechanism by which the goal-driven system influences the stimulus-driven system is still debated. The present study examined top-down contributions to two components of attention orienting, shifting and disengagement, with an experimental paradigm in which participants held a visual item in short-term memory and performed a prosaccade task with a manipulation of the gap between fixation offset and target onset. Four experiments showed that the short-term memory content accelerated shifting and impaired disengagement, but the influence on disengagement depended on the utility of short-term memory in guiding attention toward the target. Thus, the use of short-term memory was strategic. Computational models of visual attention were fitted to the experimental data, which suggested that the top-down contributions to shifting was more prominent than those to disengagement. The present study shows that the current modeling framework was particularly useful when examining the contributions of theoretical constructs for the control of visual attention, but it also suggests limitations.

Keywords: Spatial attention; attention capture; disengagement; memory-guided attention; gap task.

The amount of visual information that can be processed at a given moment is limited, which imposes a challenge for the cognitive system to select a portion of the visual information for further analyses. This selection process depends on spatial attention. The orienting of spatial attention can be divided into three major steps (Posner, 1980): disengagement of attention from the current attentional focus, shift of the attentional focus toward a new location, and engagement of the attentional focus to the new location. The majority of theories of visual attention implicate two separate systems to control these operations (Itti & Koch, 2000; Jonides, 1981; Müller & Rabbitt, 1989; Posner, 1980; Theeuwes, 1991). The first is an endogenous system that controls spatial attention according to a specific goal of a task (e.g., finding a green wagon in a car park). The second is an exogenous system that is driven by external informational factors such as the saliency of visual objects. The exogenous system is thought to work independently of the endogenous system, and different principles have been suggested to underlie the operations of the two attentional control systems (e.g., Yantis & Jonides, 1990). Nevertheless, several lines of studies have provided evidence indicating that the goal-driven process can influence the stimulus-driven process (Bacon & Egeth, 1994; Chun & Jiang, 1998; Downing, 2000; Folk, Remington, & Johnston, 1992; Found & Müller, 1996), and the mechanism by which the two systems interact is still a subject of active debate (Folk, & Remington, 2010; Gaspelin, Ruthruff, & Lien, 2016; Theeuwes, 2010). The present study addresses this issue, focusing particularly on the contributions of the top-down process to two components of stimulus-driven attention orienting, shifting and disengagement.

The present article first reports a series of four experiments that used a novel method to examine the contributions of the top-down process to shifting and disengagement of spatial attention. We then develop computational models of attention control to examine the role of the top-down process in stimulus-driven attention capture. These models are also used to test two competing theories that are aimed to explain the top-down influence on attention orienting. The results of the present study demonstrate that top-down processes affect both shifting and disengagement of attention in a strategic manner, and the contingency between attention orienting and the task goal relies mainly on the utility of top-down processes in guiding attention toward the target.

Bottom-Up and Top-Down Control of Spatial Attention

Researchers have argued that spatial attention depends on two distinct systems (Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Jonides, 1981; Posner, 1980; Yantis, 1998). The exogenous system produces stimulus-driven orienting of attention to the location or object with high saliency. For instance, when reacting to a pre-specified target, a small dot facilitates responding to the target if it occurs at the target location prior to the target onset, implying that attention has been captured by the dot at the target location. This form of attention orienting is thought to occur automatically (Itti & Koch, 2000; Theeuwes, 2010). The endogenous system produces goal-driven orienting of attention to the location at which the target is expected to occur. The system relies on the knowledge of the target or prior experiences that allow the individual to intentionally prioritize processing of information at a particular location or object. Neuroimaging studies suggest that these two systems correlate with distinct brain networks (a right ventral frontoparietal network for the exogenous system and a dorsal frontoparietal network for the endogenous system; Corbetta & Shulman, 2002), supporting the distinct roles of bottom-up and top-down orienting of attention.

Theories of visual attention have divided two stages of visual processing (Cave & Wolfe, 1990; Treisman & Gelade, 1980). The first, preattentive stage decomposes primitive features of a visual scene (e.g., color, orientation, form, etc.) and analyzes them in separate *feature maps*. These feature maps analyze different features of the visual scene in parallel before attention is applied to them. The preattentive stage may be sufficient to detect the presence of a pre-specified target if it is distinct from other objects in the visual scene (e.g., a red target among green distractors). In such a display, non-target distractors are distinct from the target along the dimension that corresponds to one of the feature maps, and the target "pops out" of the context, which allows attention to be directed to the target without attentive visual processing. However, if the target is not unique from distractors in a single dimension (e.g., a red square target among red triangles and green squares), then the results of analysis from different feature maps have to be integrated at the second stage. This integration requires attention, and different objects in the visual scene have to be processed separately in serial. The search at the second stage is guided by the first stage in which the activations of different feature maps are aggregated to create a *saliency map* (Borji & Itti, 2013; Cave & Wolfe, 1990; Itti & Koch, 2000; Oliva & Torralba, 2001). Focused attention is navigated on the saliency map endogenously.

Duncan and Humphreys (1989) proposed an alternative theory of visual attention in which the visual scene is analyzed into segments (or objects) rather than features (also see Bundesen, 1990), according to the Gestalt grouping principles (particularly the principle of similarity). This segmentation is a preattentive process and occurs throughout the visual scene in parallel. These segments compete for access to visual short-term memory that is limited in attentional capacity, and one of the segments is selected for short-term memory access if it matches a *template* maintained in the short-term memory. This is the basis of a more recent neurocognitive theory, known as the *biased competition theory* (Desimone & Duncan, 1995), in which the memory template biases the selection of a segment for visual short-term memory. This top-down bias from a memory template may modulate the activities of the ventral network implicated for the exogenous system (Corbetta, Patel, & Shulman, 2008; Kanwisher & Wojciulik, 2000; McElree & Carrasco, 2001).

Consistent with the biased competition view, a number of studies have demonstrated the dependency of stimulus-driven attention capture on the task goal or target. Bacon and Egeth (1994) suggested that bottom-up attention capture depends on the mode of search, and it occurred only when their subjects searched for a singleton (defined as an object that is distinct from other objects in any of the features) but not when they searched for a specific target feature. Yantis and Egeth (1999) found that a salient distractor did not capture attention when it did not coincide with the target that was being searched for. Similarly, Folk et al. (1992) also showed that peripheral cues captured attention only if they contained a target feature to be searched for. In their contingent cuing paradigm, a variation of the peripheral cuing paradigm, participants respond to a visual target defined by a specific feature (e.g., red circle), whereby a peripheral cue precedes the target by 150 ms at one of the potential target locations. When the peripheral cue contains the target-defining feature (red), responses are faster if the cue appears at the same location as the target (valid trials) than if the peripheral cue appears at a different location (invalid trials). However, when the peripheral cue does not contain the target-defining feature, this cue validity effect is very small or non-existent.

Furthermore, in another variation of the peripheral cuing task (Downing, 2000), participants maintain a visual stimulus (e.g., a face) in short-term memory for a later recognition memory task administered at the end of each trial. During the retention interval, two irrelevant cues are presented briefly, one that is the same as the memory item and the other that is novel. The cues are followed by the target to which participants have to respond as quickly as possible. Responses are faster when the target appears at the location that is occupied previously by the face that is identical with the memory item than when it appears at a location that is occupied by a novel face. This finding is also consistent with the biased competition theory in that attention capture depends on the content of short-term memory, even though the content is irrelevant to responding to the target.

Mechanisms Underlying Top-Down Influences on Stimulus-Driven Attention Capture

Although the exogenous and endogenous systems operate differently, stimulus-driven attention orienting is still subject to top-down control. Nevertheless, the mechanism underlying the top-down influence on attention capture is still unclear. The *contingent attention account* proposes that attention capture is not purely stimulus-driven but is contingent on an *attentional set* that is tuned to the target-defining feature in a top-down manner (Folk et al., 1992). An attentional set may be created by changing the weights of the feature maps that correspond to the target-defining properties of stimuli. This would allow attention to be deployed quickly to an object that contains target-defining features. The biased competition theory proposes that an object that matches a short-term memory template is prioritized for access to short-term memory, thus facilitating deployment of spatial attention to that object (Desimone & Duncan, 1995; Downing, 2000; see also Lu & Dosher, 1998, for a similar idea).

Yet, some researchers still maintain that attention capture is purely stimulus-driven (Theeuwes, 2010). This position holds that any stimulus that contains a salient feature (i.e., singleton) can capture attention, but it takes longer to disengage attention from the singleton when the singleton contains a target-defining feature than when it does not. Thus, top-down processes do not affect a shift of attention toward a target-defining feature but do affect disengagement from such a feature that occurs only after attention has been shifted to a cued location. Although it has been shown in the contingent cuing paradigm that there is little effect of a peripheral cue that does not match the target-defining feature, the purely stimulus-driven account predicts that the cuing effect can be observed if the interval between the cue and the target (stimulus onset asynchrony, or SOA) is shorter than the time it takes to disengage from the cued location. This prediction was confirmed in a singleton search task (Theeuwes, Atchley, & Kramer, 2000), in which participants looked for a target among distractors, one of which contained a feature distinct from other distractors (e.g., red distractor among green distractors). Although the singleton did not contain a target-defining feature, it still captured attention when the SOA between the singleton and the target was shorter than 100 ms, but it no longer captured attention when the SOA was longer than 150 ms, indicating that attention was disengaged from the singleton quickly when the singleton did not match the target. Curiously, however, other studies using the contingent cuing paradigm failed to find attention capture by the peripheral cue that did not contain a target-defining feature even when the SOA was as short as 35 ms (Chen & Mordkoff, 2007; also see Anderson & Folk, 2012; Folk & Remington, 2006).

More recently, Gaspelin et al. (2016) suggested that the abrupt onset of a peripheral cue always captures attention, consistent with the purely stimulus-driven account, but the size of the cuing effect depends on the difficulty of the search task, with easier search conditions producing smaller cuing effects. They proposed that attention ‘dwells’ at the cued location even after the cue has been erased. When visual search is easy, the distractor at the cued location can be rejected quickly, producing a small or no cuing effect. When visual search is difficult, rejection of the distractor takes time, producing a large cuing effect. They noted that search in the contingent cuing paradigm is usually easy, whereas search in the singleton search task is difficult, which explains the discrepancy between the two paradigms. The dwelling account suggests that the lack of the cuing effect in the contingent cuing paradigm is due to rapid disengagement from the distractor items (as opposed to rapid disengagement from the peripheral cue that guided attention to the distractor location). It is not explicit in the dwelling account as to how the top-down process influences attention capture, but it seems to imply that disengagement is impaired if the distractor contains a target-defining feature, consistent with the purely stimulus-driven account. However, it still seems necessary for the dwelling account to assume a top-down influence on attention shifting in order to explain a typical outcome in the contingent attention task, in that the distractor in that task does not contain a target-defining feature.

Part of the difficulty assessing the top-down influences on attention capture is attributable to the fact that it is unclear how short the SOA has to be in order to observe attention capture (Folk & Remington, 2010). Another reason is that previous studies have not been able to disentangle the top-down contributions to different components of attention orienting. As the contingent attention account and the purely stimulus-driven account disagree as to the component that the top-down control influences, it is necessary to dissociate the influences of the goal-driven process on different components of orienting, notably, those on shifting and disengagement. In the present study, we developed a paradigm to examine the top-down influences on shifting and disengagement of spatial attention.

Furthermore, we also developed a template model of visual attention and fit it to the response time (RT) distributions to account for the major findings of the experiments. The model was used to derive parameters that correspond to different cognitive constructs, such as shifting and disengagement operations of attention, and examine whether these constructs had any influences on task performance. The model framework was also used to examine competing theories of top-down influences on stimulus-driven attention capture, whereby fits of computational models that represented the competing theories were compared to examine which of these theories explained the experimental data better. The results of the experiments and computational modeling provided detailed analyses into which components of attention orienting are responsible for the top-down influences on attention capture.

Experiments

The present experiments involved a prosaccade task in which participants look toward a visual target presented to the left or right following a central fixation. In a similar task, Saslow (1967) demonstrated that saccadic responses are faster when there is a temporal gap between the fixation offset and the target onset (*gap trial*) than when the fixation mark remains on after the target onset (*overlap trial*). This is known as the *gap effect*. Arguably, a major component of the gap effect represents disengagement of spatial attention (Fischer & Breitmeyer, 1987; Mackeben & Nakayama, 1993; Pratt, Lajonchere, & Abrams, 2006; see also Jin & Reeves, 2009; Kingstone & Klein, 1993; Pratt, Bekkering, & Leung, 2000). That is, the fixation offset allows pre-disengagement of attention from the fixation mark prior to the target onset, saving the time to disengage from the fixation after the target onset. Whereas the peripheral cuing task requires three steps to make a response (shifting to the peripheral cue, disengaging from the cue, and shifting to the target), the gap task only requires two steps on overlap trials (disengaging from the fixation and shifting to the target) and one step on gap trials (shifting to the target). Thus, it disentangles the influences of top-down processes on shifting and disengagement of attention orienting.

The four experiments examined the separate contributions of top-down processes on shifting and disengagement of attention orienting (Posner, 1980) by manipulating the properties of the target and fixation. To assess the top-down contribution to attention orienting, participants performed a short-term memory task that required maintaining a memory item in short-term memory for recall, and they performed the prosaccade task during the retention interval. Downing (2000) has shown that attention is captured by a visual stimulus that matches the memory item held in short-term memory even though the short-term memory content is irrelevant to the goal of participants at the moment. Experiment 1 tested whether the short-term memory content affected shifting of spatial attention, and Experiment 2 tested whether the short-term memory content affected disengagement of spatial attention. Experiments 3 and 4 further tested the role of strategy in the top-down contributions to shifting and disengagement.

General Method

Participants

One hundred and twenty participants were recruited for the present study from the same subject pool consisting of students and staff members at Edge Hill University. Participants were given experimental credits toward their psychology module or paid £3-6, depending on the length of their sessions. All reported having normal or corrected-to-normal visual acuity, normal color vision, and normal hearing. All were naïve as to the purpose of the experiment. The experimental procedure was approved and monitored by the Research Ethics Committee of the Department of Psychology at Edge Hill University.

Moderate to large effect sizes were expected in the present experiments. With a within-subject design and the statistical power of .8, the appropriate sample size ranges between 10 and 20, depending on the number of levels in the given variable.

Twenty four participants were assigned to each of the first three experiments, and forty eight participants were assigned to the last experiment. There were 18 females and 6 males in Experiment 1 (mean age = 20.58, *SD* = 4.19, range = 18-37) and Experiment 2 (mean age = 19.88, *SD* = 2.44, range = 18-26), 22 females and 2 males in Experiment 3 (mean age = 18.79, *SD* = 1.87, range = 18-26), and 32 females,15 males, and 1 undeclared, in Experiment 4 (mean age = 20.81, *SD* = 2.86, range = 18-36).

Apparatus and Stimuli

The same apparatus and stimuli were used in the four experiments. The apparatus consisted of a personal computer with a 19-in. CRT monitor. The experiment was controlled by an E-Prime 2.0 program (Psychology Software Tools, Pittsburgh, PA). An Eyelink 1000 (SR Research, Ontario, Canada: 1000 Hz, <0.5° accuracy) was used to monitor saccades and record RT. The memory item was a colored circle (green, red, yellow, magenta, or cyan), with a diameter of 1.2 cm (1.15°), and appeared 6.5 cm (6.20°) above the screen center. The fixation mark was a circle (1.2 cm in diameter; 1.15°) that appeared in the center of screen. The target was a square with the side length of 1.2 cm and appeared 11.2 cm (10.66°) to the left or right of the screen center.

Task

The experiment was conducted individually in a cubicle under normal fluorescent lighting. Participants were seated in front of the computer monitor at a distance of 60 cm from the monitor and rested on a chinrest throughout the session. A standard 9-point calibration and validation procedure was employed to calibrate the eye tracker at the beginning of the session. Experiments 1 and 2 involved single- and dual-task conditions, and Experiments 3 and 4 only involved the dual-task condition. The sequence of events on a trial is depicted in Figure 1.

In the single-task condition, each trial started with a crosshair at the center of screen. Participants fixated on the cross and pressed the space bar. This served as the drift correction, and the eye tracker started recording as participants pressed the space bar. Then, the fixation circle appeared at the center of screen for variable durations (800, 900, or 1000 ms) and was replaced by a 200-ms blank display on gap trials or remained on the screen until the end of the trial on overlap trials. A high-pitch tone was presented 200 ms before the target onset as a warning signal[1]. The target was a square that appeared on the left or right side of the screen. Participants moved their gaze to the target as quickly and as accurately as they could. The target lasted for 1000 ms. RT was defined as the interval between the target

---

[1] For a technical reason, the warning tone was not presented to some participants in Experiments 1 and 2, but the data of these participants were not different from those to which the tones were presented correctly. Also, the main results (gap effect) of Experiments 1 and 2 were replicated in Experiments 3 and 4 in which the warning tone occurred for every participant.

onset and the first saccade greater than 2° to the left or right. A response was considered an error if the initial saccade was made to the opposite side of the target position.

In the dual-task condition, the task was similar to the single-task condition, except that participants also performed the memory task. A trial started with the memory item that occurred 6.5 cm above the screen center along with the crosshair at the center. Participants fixated the cross and pressed the space bar, performing the drift correction. The fixation circle appeared at the screen center and stayed on the screen on overlap trials or disappeared 200 ms before target onset on gap trials. As 1000 ms elapsed after the target onset, five diamond shapes appeared around the target with the equal angular distance between any of two adjacent diamonds. These diamonds had five unique colors. Participants had 2000 ms to indicate the color of the memory item by moving the mouse cursor to one of the diamonds. The positions of the colors in the diamonds were determined randomly on each trial. As the mouse cursor touched one of the diamonds, a low-pitch tone was presented. Feedback was provided by the message “Correct!” if the correct color was indicated, “Error!” if a wrong color was indicated, and “Faster!” if no response was made within 2000 ms.

Experiment 1

Experiment 1 focused on the top-down contribution to attention shifting. To examine the contribution of top-down processes, the content of short-term memory was manipulated in the dual-task condition. The target for the prosaccade task either matched or did not match the color of the memory item. If the content of short-term memory affects shifting of attention, responses should be faster when the target color matches the color of the memory item than when the target color does not match. If top-down processes do not affect shifting of spatial attention, saccade responses should be as fast when the target matches the color of the memory item as when it did not match. Also, it is possible that top-down processes affect a disengagement operation, albeit indirectly, by facilitating attention shifting. If so, the gap effect (i.e., faster responses when there is a temporal gap between fixation offset and target onset than when the fixation stays on throughout a trial) should be smaller when the target color matches the color of the memory item than when it does not match (note that top-down contribution to disengagement was examined more directly in Experiment 2).

Procedure

A session started with two practice blocks. The first practice block consisted of the single-task condition. Half the trials were gap trials, and the other half were overlap trials. There were a total of 8 trials in this block. The second practice block consisted of the dual-task condition, and there were 12 trials. After the practice blocks, participants performed three test blocks of 72 trials each. One test block was the single-task condition, and the remaining two blocks were the dual-task condition. The order of the test blocks was determined randomly for each participant. In the single-task condition, the fixation was always colored in white, the target was always colored in gray. In the dual-task condition, the color of the target matched the color of the memory item on half of the trials (*target match*), and it did not match on the other half (*target mismatch*). On latter trials, the color was chosen randomly from the four remaining colors. The fixation was always in white in this condition.

Results and Discussion

Mean RTs were computed for each participant, including trials for which the color of the memory item was identified correctly (mean error percentage = 4.08%). Trials were excluded if the amplitude of the first saccade after target onset was less than 2° or if the latency of the first saccade was less than 80 ms or greater than 800 ms (3.42% of all trials were discarded). RTs are summarized in Figure 2a and were analyzed separately for the single-task and dual-task conditions. For the single-task condition, RTs for gap and overlap trials were submitted to a paired-sample *t*-test, which revealed significantly faster responses on gap trials (*M* = 132 ms) than on overlap trials (*M* = 177 ms), $t(23) = 7.17$, $p < .001$, $d = 1.464$. RTs for the dual-task condition were submitted to a 2 (Target Match: match vs. mismatch) x 2 (Trial Type: gap vs. overlap) ANOVA. There was a main effect of Target Match, $F(1, 23) = 15.27$, $MSE = 191.03$, $p = .001$, $\eta_p^2 = .399$. RT was shortest for target match trials (*M* = 177 ms) than for target mismatch trials (*M* = 188 ms), which indicates that attention shifting was faster when the target color was the same as the color of the memory item than when it was different. There was also a main effect of Trial Type, $F(1, 23) = 61.12$, $MSE = 539.08$, $p < .001$, $\eta_p^2 = .727$. RT was shorter for gap trials, (*M* = 164 ms) than for overlap trials (*M* = 201 ms), yielding a gap effect. The interaction between the two factors did not reach the significance level, $F(1, 23) = 3.05$, $MSE = 338.23$, $p = .094$, $\eta_p^2 = .117$. Overall, the results indicated that top-down processes enhanced shifting of attention toward the target that matched the color of a memory template. The gap effect did not depend on the match between the target color and the memory item, indicating little top-down influence on attention disengagement, but Experiment 2 provided a more direct test on this issue.

Experiment 2

Experiment 2 tested the top-down contribution to attention disengagement. Unlike Experiment 1, the target was always colored in gray. This condition excluded the top-down influence on attention shifting observed in Experiment 1. Instead, the present experiment isolated the top-down contribution to disengagement. The color of the fixation mark matched that of the memory item on half of the trials (*fixation match*) but did not match on other half (*fixation mismatch*). If top-down processes affect disengagement, it should take longer to remove attention away from the fixation mark when the fixation color matches the content of short-term memory. However, such slowing should occur only on overlap trials, but not on gap trials that allowed pre-disengagement of attention regardless of whether the fixation color matches the short-term memory content. Therefore, it was expected that the gap effect would increase when the fixation color matched the color of the study item, as compared to when the fixation color differed from the short-term memory content. If top-down processes had no effect on disengagement, the gap effect should not depend on whether the fixation color matched the color of the memory item.

Procedure

Experiment 2 only differed from Experiment 1 in that the color of the fixation was varied in the dual-task condition. Half of the trials in the test blocks were fixation match trials for which the color of the fixation matched the color of the memory item, and the other half were fixation mismatch trials for which the color of the fixation did not match the color of the memory item. The target appeared in gray on all trials.

Results and Discussion

Trials were filtered in the same manner as in Experiment 1. One of the participants was excluded because more than half of the trials were discarded. Among the remaining 23 participants, 4.86% of all trials were discarded. Mean RT was computed for each participant, including trials with correct response to the memory task (mean error rate = 4.86%). The results are summarized in Figure 2b.

For the single-task condition, RT was significantly shorter for gap trials (M = 141 ms) than for overlap trials (M = 175 ms), $t(22) = 5.96$, $p < .001$, $d = 1.242$. For the dual-task condition, RT was submitted to a 2 (Fixation Match: match vs. mismatch) x 2 (Trial Type: gap vs. overlap) ANOVA, which revealed a main effect of Trial Type, $F(1, 22) = 35.78$, $MSE = 845.33$, $p < .001$, $\eta_p^2 = .619$. RT was shorter for gap trials (M = 160 ms) than for overlap trials (M = 197 ms). There was no significant main effect of Fixation Match, $F(1, 22) < 1$, $MSE = 363.03$, $p = .957$, $\eta_p^2 < .001$, or its interaction with Trial Type, $F(1, 22) = 2.17$, $MSE = 155.84$, $p = .155$, $\eta_p^2 = 090$. Thus, the results provided little evidence that the short-term memory content had any influence on disengagement from the fixation. This outcome is in contrast to the top-down influence on shifting observed in Experiment 1.

Experiment 3

Although Experiment 2 did not provide evidence supporting any top-down influence on disengagement, it is possible that the outcomes were due to a strategic choice of participants who decided not to use the short-term memory template to guide attention (Woodman & Luck, 2007). In that experiment, participants always had to shift their gaze away from the fixation to the target. As the fixation was the only possible stimulus that could match the short-term memory content in Experiment 2, short-term memory would only slow down the shift toward the target but would never facilitate it. As studies have suggested, the use of short-term memory content could be strategic rather than strictly automatic (Arita, Carlisle, & Woodman, 2012; Carlisle & Woodman, 2011; Woodman & Luck, 2007). Hence, it is possible that participants decided to not use a memory template in short-term memory to guide their attention. To examine this possibility, Experiment 3 tested the top-down influence on attention disengagement under the condition in which a memory template was actually used to guide visual attention toward the target. This was accomplished by manipulating both the match/mismatch of the target and fixation colors with the memory item.

As in Experiment 1, the target color matched the study item color on half of the trials, but it did not match on the other half. The top-down contribution to attention shifting was observed in these conditions, so we expected that participants would make use of the short-term memory content to guide their attention in the present experiment if the use of short-term memory content is strategic. Furthermore, the fixation color also matched the study item color on half of the trials, which was manipulated orthogonally to the target color. We expected that if participants make use of the short-term memory content to guide their attention, the short-term memory content should affect both shifting toward the target and disengagement from the fixation in the present experiment. Shifting toward the target should be faster when the target color matches the study item color than when it did not match. Disengagement from the fixation should be slower, and the gap effect should be larger, when the fixation color matches the study item color than when it did not match. If the short-term memory content only affects attention shifting, then the gap effect should not depend on whether the fixation color matches the study item color as in Experiment 2.

Procedure

Experiment 3 started with a block of 16 practice trials, followed by three blocks of 80 test trials. All blocks were the dual-task condition in which both the fixation color and target color were manipulated to match or mismatch the memory item. The target color matched the color of the memory item on half of the trials (*target match*), and it did not match on the other half (*target mismatch*). In each of these conditions, the fixation color matched the color of the memory item on half of the trials (*fixation match*), and it did not match on the other half (*fixation mismatch*). All combinations of target and fixation match/mismatch occurred equally frequently in a random order.

Results and Discussion

Trials were filtered in the same manner as the preceding experiments (3.57% of all trials were discarded), and mean RT was computed for each participant. Mean error rate for the memory task was 2.90%. Mean RTs for the prosaccade task are summarized in Figure 2c. RTs were submitted to a 2 (Target Match: match vs. mismatch) x 2 (Fixation Match: match vs. mismatch) x 2 (Trial Type: gap vs. overlap) ANOVA.

There was a significant main effect of Target Match, $F(1, 23) = 25.30$, $MSE = 270.87$, $p < .001$, $\eta_p^2 = .524$. RT was shorter when the target color matched with the study item color ($M = 162$ ms) than when it did not ($M = 174$ ms). This reproduced the top-down influences on attention shifting that was observed in Experiment 1. There was also a main effect of

Trial Type, *F*(1, 23) = 82.54, *MSE* = 382.93, $p < .001$, $\eta_p^2 = .782$, yielding shorter RT for gap trials (*M* = 155 ms) than for overlap trials (*M* = 181 ms). This variable did not interact with Target Match, *F*(1, 23) = 2.72, *MSE* = 93.70, $p = .112$, $\eta_p^2 = .106$. Thus, there was little evidence that the target color affected disengagement from the fixation mark. To this point, the results confirmed that participants used short-term memory to guide attention toward the target, which sets a precondition to test the top-down influence on disengagement.

There was no significant main effect of Fixation Match, *F*(1, 24) < 1, *MSE* = 171.78, $p = .340$, $\eta_p^2 = .040$, but the variable interacted with Trial Type, *F*(1, 23) = 9.40, *MSE* = 129.29, $p = .005$, $\eta_p^2 = .290$. The gap effect was larger when the fixation color matched the study item color (*M* = 31 ms) than when it did not (*M* = 21 ms). The outcomes indicate that it took longer to disengage attention from the fixation mark when the short-term memory content was identical with the fixation color than when it differed. Therefore, top-down processes do impair attention disengagement when short-term memory is used to guide attention toward the target. The three-way interaction among Trial Type, Target Match, and Fixation Match, was not significant, *F*(1, 23) < 1, *MSE* = .158.09, $p = .629$, $\eta_p^2 = .010$.

Experiment 4

To follow up the role of strategic control implied in Experiment 3, the present experiment varied the proportions of target match and target mismatch trials across blocks. In one block of trials (*mostly target match*), the colors of the target and memory item matched on 80% of the trials, and they mismatched on 20% of the trials. In the other block (*mostly target mismatch*), the colors of the target and the memory item matched on 20% of the trials, and they mismatched on 80% of the trials. The colors of the fixation and the memory item matched on 50% of the trials in each of these conditions. It was expected that participants would make use of the short-term memory template to facilitate shifting toward the target if the target color matched the memory item in most of the trials. Thus, in the mostly target match block, participants would responded to the target faster on target match trials than on target mismatch trials. Nevertheless, the use of the short-term template would impair disengagement from the fixation, so the gap effect would be larger on fixation match trials than on fixation mismatch trials in this block. On the other hand, participants would not make use of the short-term memory template if the target color mismatched the memory item in most of the trials. Thus, in mostly target mismatch block, participants would respond to the target equally fast on target match and target mismatch trials, and the gap effect would not depend on fixation match or mismatch.

Procedure

Experiment 4 consisted of two phases, each phase being essentially the same as Experiment 3 but with varying proportions of target match and target mismatch trials. In the mostly target match block, the target color was the same as the memory item color on 80% of the trials, and it differed from the memory item color on 20% of the trials. In the mostly target mismatch block, the target color was the same as the memory item color on 20% of the trials, and it differed from the memory item color on 80% of the trials. The fixation color was the same as the memory item color on 50% of the trials, and it differed on another 50% of the trials, in each of these conditions. Participants were not informed of the proportions of target match and target mismatch trials prior to the task. As the strategy of participants might depend on the order of the conditions, two groups of 24 participants were recruited, who performed the two conditions in different orders: one group performed the mostly target match condition and then the mostly target mismatch condition; the other group performed the mostly target mismatch condition and then the mostly target match condition. Each phase consisted of one block of 16 trials and three blocks of 80 trials each, and started with the same instructions. Each of the two phases was the same duration as that of Experiment 3, so the overall duration of a session was doubled.

Results and Discussion

Trials were filtered in the same manner as the preceding experiments. Three participants ended up losing more than 30% of their trials, and these participants were excluded from the analysis. This left 23 participants who started with the mostly target match condition, and 22 participants who started with the mostly target mismatch condition. For the remaining participants, 5.55% of all trials were discarded. Mean error rate for the memory task was 2.28%. Mean RTs for the prosaccade task are summarized in Figure 3.

RTs were first submitted to a 2 (Target Match: match vs. mismatch) x 2 (Fixation Match: match vs. mismatch) x 2 (Trial Type: gap vs. overlap) x 2 (Trial Proportion: mostly target match vs. mostly target mismatch) x 2 (Order: mostly target match first vs. mostly target mismatch first) mixed-design ANOVA. The first four variables were within-subject factors, and the last variable was a between-subject factor.

There were main effects of Target Match, $F(1, 43) = 16.95$, $MSE = 338.58$, $p < .001$, $\eta_p^2 = .283$, Fixation Match, $F(1, 43) = 4.77$, $MSE = 267.80$, $p = .034$, $\eta_p^2 = .100$, Trial Type, $F(1, 43) = 35.14$, $MSE = 612.59$, $p < .001$, $\eta_p^2 = .450$, and Trial Proportion, $F(1, 43) = 10.43$, $MSE = 141.85$, $p = .002$, $\eta_p^2 = .195$. Trial Proportion modulated Target Match, $F(1, 43) = 12.55$, $MSE = 564.11$, $p = .001$, $\eta_p^2 = .226$, and Trial Type, $F(1, 43) = 43.08$, $MSE = 554.47$, $p < .001$, $\eta_p^2 = .500$. For the mostly target match condition, RT was faster when the target color matched the memory item color ($M = 155$ ms) than when the target color did not match the memory item color ($M = 167$ ms), indicating the top-down influence on attention shifting. In the mostly target mismatch condition, however, RT did not differ between target match ($M = 159$ ms) and target mismatch trials ($M = 158$ ms). Interestingly, the gap effect was obtained in the mostly target match condition (*M*s = 150 ms vs. 173 ms for gap and overlap trials, respectively) but not in the mostly target mismatch condition (*M*s = 159 ms vs. 158 ms for gap and overlap trials). This outcome seems to suggest that participants ignored the fixation mark altogether in the latter condition, presumably to avoid using the short-term memory template in guiding attention. The influence of Fixation Match was not modulated by Trial Proportion, $F(1, 43) < 1$, $p = .962$. However, there were some indications that the order of the conditions influenced the effect of Fixation Match: the interaction between Fixation Match and Order, $F(1, 43) = 3.77$, $MSE = 267.80$, $p = .059$, $\eta_p^2 = .081$; and the three-way interaction among Fixation Match, Trial Type, and Trial Proportion, $F(1, 43) = 3.52$, $MSE = 416.68$, $p = .067$, $\eta_p^2 = .076$. Both of these terms yielded medium effect sizes. These outcomes hinted that participants' strategies might have depended on which condition they performed in the first block. Therefore, the data were analyzed separately for the two groups of participants with different orders of the mostly target match condition and the mostly target mismatch condition.

Figures 3a and 3b show the results of those who performed the mostly target match condition first, and Figures 3c and 3d show the results of those who performed the mostly target mismatch condition first. RT for the two groups were submitted to a 2 (Target Match: match vs. mismatch) x 2 (Fixation Match: match vs. mismatch) x 2 (Trial Type: gap vs. overlap) x 2 (Trial Proportion: mostly target match vs. mostly target mismatch) repeated-measures ANOVAs separately. These two groups showed similar patterns of results but differed in some details. Both groups yielded a significant main effect of Target Match, $F(1, 22) = 10.36$, $MSE = 216.23$, $p = .004$, $\eta_p^2 = .320$, for those who performed the mostly target match condition first, and $F(1, 21) = 7.63$, $MSE = 466.75$, $p = .012$, $\eta_p^2 = .267$, for those who performed the mostly target mismatch condition first. There was also a main effect of Trial Type, $F(1, 22) = 17.76$, $MSE = 581.93$, $p < .001$, $\eta_p^2 = .447$, for those who performed the mostly target match first, and $F(1, 21) = 17.36$, $MSE = 644.71$, $p < .001$, $\eta_p^2 = .453$, for those who performed the target mismatch condition first. Furthermore, a main effect of Trial Proportions was significant, $F(1, 22) = 4.93$, $MSE = 111.84$, $p = .037$, $\eta_p^2 = .183$, for those who performed the target match condition first, and *F*(1, 21) = 5.48, *MSE* = 173.30, $p = .029$, $\eta_p^2 = .207$, for those who performed the target mismatch condition first. Trial Proportion modulated Target Match for both groups, *F*(1, 22) = 6.32, *MSE* = 529.11, $p = .020$, $\eta_p^2 = .223$, for those who performed the target match condition first, and *F*(1, 21) = 6.22, *MSE* = 600.79, $p = .021$, $\eta_p^2 = .229$, for those who performed the target mismatch condition first. The effect of Trial Type was also modulated by Trial Proportion, *F*(1, 22) = 23.34, *MSE* = 549.05, $p < .001$, $\eta_p^2 = .515$, for those who performed the mostly target match condition first, and, *F*(1, 21) = 19.85, *MSE* = 560.15, $p < .001$, $\eta_p^2 = .486$, for those who performed the target mismatch condition first. These interactions agreed with the outcomes of the earlier analysis.

The discrepancies between the groups were that the main effect of Fixation Match was significant only for those who first performed the mostly target match condition, *F*(1, 22) = 7.39, *MSE* = 315.61, $p = .013$, $\eta_p^2 = .251$, reflecting longer RT when the fixation color matched the memory item (*M* = 166 ms) than when it did not (*M* = 161 ms). In contrast, for those who first performed the mostly target mismatch condition, the main effect of Fixation Match was not significant, *F*(1, 21) < 1, *MSE* = 217.70, $p = .853$, $\eta_p^2 = .002$, but the variable interacted with Trial Proportion and Trial Type, *F*(1, 21) = 5.77, *MSE* = 387.41, $p = .026$, $\eta_p^2 = .215$. This interaction showed that for the mostly target match condition, the gap effect was larger when the fixation color matched the memory item (*M* = 28 ms) than when it did not (*M* = 17 ms). For the mostly target mismatch condition, however, the gap effect was rather larger when the fixation color did not match the memory item (*M* = 5 ms) than when it matched (*M* = -5 ms). The gap effects were similar for fixation match and fixation mismatch trials in the mostly target match condition (*M*s = 23 ms vs. 22 ms) and in the mostly target mismatch condition (*M*s = 1 ms vs. – 3 ms) for those who started with the mostly target match condition.

Overall, the results of the mostly target match condition are similar to the results of Experiment 3, although they depended on the order of the conditions. The gap effect increased when the fixation color matched the memory item for those who started with the mostly fixation mismatch condition (see Figure 3c), but the gap effect was not influenced by fixation match for those who started with the mostly target match condition (see Figure 3a). The outcomes of the mostly target mismatch condition suggest that there was little influence of the short-term memory template on attention shifting or disengagement for both groups (see Figures 3b and 3d). The mostly target mismatch condition produced no gap effect, which seems to suggest that participants ignored the fixation when the utility of the short-term memory template in guiding attention was lower than its cost of impairing disengagement. The gap effect was obtained in Experiment 2 in which the target always differed from the memory item and, thus, the short-term memory template had no utility in guiding attention in that experiment. The gap effect disappeared in the mostly target mismatch condition of the present experiment possibly because having target match trials in a minority of trials motivated, or even required, participants to actively “turn off” attention toward the fixation in order to avoid automatic reaction toward stimuli that matched the short-term memory template. Such automatic reactions may be triggered when there are target match trials as in the present experiment, but not when there are no target match trials as in Experiment 2.

As the present experiment involved a more complicated design than the preceding experiments, the results also showed complex patterns overall. Yet, the proportions of target match and mismatch trials were shown to be influential in modulating the effect of target and fixation match with the memory item, implying strategic control over the top-down influences on attention orienting.

Summary of the Experiments

Experiment 1 demonstrated that the content of short-term memory affected attention shifting. Participants shifted their gaze to the target more quickly when the target color was identical with the memory item than when it was different. Note that the target was the only stimulus that appeared on the screen, except for the fixation mark on overlap trials. Therefore, it should capture attention exogenously in a bottom-up manner, regardless of whether the target color matched the content of short-term memory (Gaspline et al., 2016; Theeuwes, 2000). The additional benefit of target color match with short-term memory provides an unambiguous support for top-down contribution to attention shifting that further accelerated attention shifting to the target (Reinhart, McClenahan, & Woodman, 2016).

To examine whether top-down processes influence disengagement operations, Experiment 2 assessed the influence of short-term memory content on the gap effect when the color of the fixation mark was manipulated, which excluded the possible top-down influence on attention shifting. The results provided little evidence supporting the effect of short-term memory on disengagement. Nevertheless, the use of short-term memory content could be strategic rather than strictly automatic (Arita, Carlisle, & Woodman, 2012; Carlisle & Woodman, 2011; Woodman & Luck, 2007). We noted that spatial attention always had to be shifted away from the fixation mark, so using the short-term memory content could only slow saccade responses to the target but would never facilitate them. As there was no benefit to utilize the short-term memory content to guide attention toward the target, participants might have decided not to use short-term memory. To test the strategic use of short-term memory, the top-down influence on disengagement was examined further in Experiment 3.

Experiment 3 showed that prosaccade responses were faster when the target color matched the study item color than when it did not match, which replicated the accelerated attention shifting by the short-term memory content observed in Experiment 1. Thus, even if participants could strategically choose to use short-term memory to guide attention, the outcomes indicate that they did use it in Experiment 3, and this set a strong test for the top-down influence on disengagement. The results showed that the gap effect was larger when the fixation color matched the study item color than when it did not. The larger gap effect implies that it took longer to disengage attention away from the fixation mark.

Experiment 4 corroborated the strategic use of short-term memory on attention orienting. When the target color matched the memory item on most trials, attention shifting was accelerated when the target color matched the memory item versus when it did not, and attention disengagement was impaired when the target color did not match the memory item. When the target color did not match the memory item on most trials, there was neither acceleration of attention shifting nor impairment of attention disengagement. The gap effect was also eliminated in this condition, implying that participants ignored the fixation mark when target color only matched the memory item in a small portion of trials. Taken together, the present outcomes support the conclusion that the short-term memory content can both accelerate attention shifting toward memory-matching objects and impair attention disengagement away from memory-matching objects, but the influence depends on strategic choice (Woodman & Luck, 2007).

In the remaining part, we developed computational models to examine the mechanisms underlying the top-down influences of attention orienting more formally. First, a general modeling framework was described conceptually, and we introduced a computational framework to model the top-down influence of attention orienting. Second, we tested whether the formal framework would fit to the experimental data to examine its feasibility. The resulting parameters would suggest the top-down contributions of different components of attention. Third, we extended the approach by developing model variations that represented competing theories of visual attention that propose different mechanisms of top-down influences on attention capture. We then examined which theory predicted the experimental data best.

Computational Modeling

To account for the results of the present study, we implemented a template model computationally (see Figure 4a; Bundesen, 1990; Desimone & Duncan, 1995; Duncan & Humphreys, 1989; also see Lu & Dosher, 1998; Smith & Ratcliff, 2009). The model consists of four components. The first component is an encoding system that consists of separate feature maps as typically assumed in many theories of visual attention (e.g., Itti & Koch, 2000; Treisman, 1969; Wolfe, 1994). We do not specify exactly what feature maps are available or how they represent the respective features (see Borji & Itti, 2013, for a review of various computational approaches), but we assume that activation of separate feature maps are integrated into a global saliency map that represents the entire visual field. The activation distributed across the saliency map corresponds to the likelihood of spatial attention to be deployed to the corresponding location of the visual scene in a bottom-up manner. These locations on the saliency map compete for the gaze (focal vision), and this competition is conceptualized as a race between independent accumulators toward a threshold; this accumulator process represents the second component and the main part of our computational model.

As described in more detail below, this race model is formulated in terms of diffusion accumulators (Ratcliff, 1978; also see Wolfe, 2007), and spatial attention is deployed to the location whose activation exceeds the threshold. The activation on the saliency map determines the rate of accumulation at a given location or object. In the task setting of the current study, the accumulator that represents the target location is mutually inhibitory with the accumulator that represents the fixation location (Boucher, Palmeri, Logan, & Schall, 2007). On an overlap trial in which the fixation stays after a target onset, the fixation unit interferes with the activation of the target unit, which slows saccade responses. On a gap trial in which the fixation is erased before a target onset, the activation of the target unit is free from interference from the fixation unit (assuming that 200 ms is sufficient for the fixation unit to be deactivated). Thus, the mutual inhibition between the target and fixation units naturally predicts the gap effect. After the race, a signal is then sent to the third component, the oculomotor system that makes a saccade to that location.

The fourth component is a top-down process that affects the activation of the saliency map. The activation of the saliency map depends on how weights are assigned to the feature maps (e.g., Müller & Krummenacher, 2006; Wolfe, 1994). The top-down component may affect the saliency map by adjusting the allocation of weights to the feature maps in accordance to the properties of the target. If the target is known to be red, for instance, then the feature map corresponding to the color ‘red’ would be assigned a greater weight, which then enhances the activation of the location on the saliency map that is occupied by red objects. The manipulation of short-term memory content can be modelled in a similar manner. The feature maps that are consistent with the properties of a memory template in short-term memory may be assigned greater weights, or it is equally possible that the short-term memory directly affects the activation of the global saliency map, in a way that the location occupied by an object that matches the memory template is enhanced (Desimone & Duncan, 1995). Note, however, that although we hypothesize this component to involve visual short-term memory as depicted in Figure 4a, the design of the present experiments do not rule out a possibility that stimulus colors were represented verbally because they were categorically unambiguous and could be named easily. The experiments did not prevent verbal encoding by requiring articulatory suppression either. Therefore, the memory template might have reflected non-visual representations. In either case, the result is that the location on the saliency map that contains an object matching the features of the memory template is activated more strongly, and spatial attention is more likely to be deployed to that location. If the location happens to contain the target, saccade responses are facilitated, resulting in faster shifting of attention. If the location happens to contain the fixation, there will be stronger inhibition from the fixation unit to the target unit, impairing disengagement from the fixation.

The proposed model can be implemented computationally. For instance, the underlying process of visual information accrual can be instantiated within the framework of the *interactive race model* (see Figure 4b) that was originally developed for saccade countermanding (Boucher et al., 2007), which was refined recently to include detailed dynamics of the inhibitory process (Logan, Yamaguchi, Schall, & Palmeri, 2015). The interactive race model consists of two competing accumulator units, *go* and *stop* (see the second component in Figure 4a). The go unit represents the activation of a saccade response that occurs when the activation of the go unit reaches a certain threshold value. The stop unit represents the inhibitory control on the go unit, suppressing a saccade response until the go unit accrues sufficient evidence indicating the presence of the target. The strength of the inhibition is proportional to the level of activation of the stop unit, and the level of activation of the stop unit depends on the input from the fixation mark as it has been assumed in the previous studies (Boucher et al., 2007; Logan et al., 2015).

Applying the interactive race model to the present task settings with the gap manipulation (see Figure 5), there are differences in the dynamics of the stop unit activation between overlap and gap trials, which then affect the dynamics of the go unit activation by way of lateral inhibition. On a gap trial, the fixation mark is erased 200 ms prior to the target onset. This allows the stop unit to be deactivated quickly and the go unit activation to raise rapidly upon a target onset. On an overlap trial, the fixation mark stays on the screen until the trial ends. The stop unit activation stays at a high level and continues inhibiting the go unit activation after the target onset, which slows the completion of the go unit process. Therefore, the model predicts a slower activation rate on overlap trials than on gap trials, hence producing the gap effect. Top-down process can be considered to be an amplification of the activation of the go or stop unit, depending on whether the short-term memory content matches the target or the fixation. If it matches the target, the activation rate of the go unit is facilitated, and saccade response can be produced quickly; if it matches the fixation, the activation of the stop unit is facilitated, and the stop unit inhibits the go unit more strongly, making it difficult to disengage from the fixation.

The interactive race model has been formalized within the framework of Usher and McClelland's (2001) leaky competing accumulator (LCA) model, which is defined for the go unit by the following stochastic differential equation:

$$da_{go} = \frac{dt}{\tau}\left[v_{go} - \kappa a_{go}(t) - \beta_{stop} a_{stop}(t)\right] + \sqrt{\frac{dt}{\tau}}\xi, \qquad (1)$$

where a_{go} and a_{stop} are the activation levels of the go and stop units, v is the drift rate, κ is the decay parameter proportional to the activation level of the go unit, and β is the strength of lateral inhibition proportional to the activation level of the stop unit. The equivalent equation can be derived for the stop unit by replacing the subscriptions 'go' and 'stop' (see Appendix A). As it does not have a closed mathematical form, the implementation of the interactive race model within the LCA formalization requires simulating each iteration of the process to derive its predictions. This is computationally very expensive. The LCA formalization is most useful when the dynamics of the accumulation process is of the main concern, which may be the case, for example, if main research questions concern comparisons of the model dynamics to neural activities (Boucher et al., 2007; Logan et al., 2015). Nevertheless, a detailed dynamics of the underlying accumulation process is not particularly important or consequential when the main interest rests in the analysis of behaviour (e.g., Brown & Heathcote, 2005, 2008; Jones & Dzhafarov, 2014). Therefore, we decided to utilize an alternative implementation of the interactive race model that is as powerful as the LCA formalization but is computationally less expensive and mathematically more tractable. For those interested, we also present a comparison between a former interactive race model (Logan et al., 2015) and the present implementation of the model in Appendix A.

Model Specifications

At the conceptual level, the present model adopts the assumptions of the interactive race model. As in the LCA framework, we considered the activation of the go unit as a single accumulator diffusion process (see Figure 4c) with a fixed rate of activation (drift rate, v). The drift rate is the mean accumulation rate of the process that determines how quickly the process completes. The larger the drift rate is, the faster the process completes. This accumulation process can be represented by a well-known Wiener process, whose state at the time t is expressed by $X(t) = vt + N(0, \sigma)$, where $N(0, \sigma)$ stands for a random error component that is distributed normally with zero-mean and variance σ. Setting a specific boundary value, or threshold z, the time t at which the state exceeds the threshold (i.e., $X(t) > z$) for the first time after an onset of the process (i.e., first passage time) is given by a random variable that has an inverse Gaussian, or Wald, distribution (Logan, Van Zandt, Verbruggen, & Wagenmaker, 2014). This Wald diffusion model has the probability density function (PDF):

$$f(t) = z[2\pi\sigma^2 t^3]^{-\frac{1}{2}}\exp\{-\frac{1}{2\sigma^2 t}[vt - z]^2\}, \qquad (2)$$

and the cumulative distribution function (CDF):

$$F(t) = \Phi\left[\left(\frac{1}{\sigma^2 t}\right)^{\frac{1}{2}}(vt - z)\right] + \exp\left\{\frac{2vz}{\sigma^2}\right\}\Phi\left[\left(\frac{1}{\sigma^2 t}\right)^{\frac{1}{2}}(vt + z)\right] \qquad (3)$$

where Φ is the CDF of the standard normal variable.

In our implementation of the model, the activation of the go unit starts as soon as the target appears on the monitor, and the process completes and a saccade response is executed at the moment the activation level reaches the threshold. The completion time of the go unit reflects the *decision time* T_D. RT also includes the non-decision time t_0 that reflects all other processes outside the go activation (e.g., encoding of stimuli, planning and execution of the selected gaze movement). Thus, we defined $RT = T_D + t_0$. The non-decision time can be incorporated into Equations 2 and 3 by substituting t with $RT – t_0$. The error variance σ is only a scale parameter, and we set it to the unity. Because the present experiments used simple prosaccade tasks and involved no choice alternative, Equation 2 alone was sufficient to represent the task performance.

There were five parameters (a, i, z, t_0, and r) for the dual-task conditions of the present study. The activation a determined how quickly the process accumulates toward the threshold z, and it is considered to reflect a bottom-up process that is based on the activation on the saliency map in which the target is represented. On an overlap trial, the fixation mark interferes with the activation from the target, which is expressed by the inhibition i. Thus, when the target and the fixation are present simultaneously, the drift rate is given by $v = a – i$. On a gap trial, the fixation is no longer present after the target onset, so there is no inhibition ($i = 0$) and the drift rate is given by $v = a$. Consequently, the difference between the overlap and gap trials is i, which is responsible for the gap effect. If the short-term memory content is used to guide visual attention, then the top-down influence is represented by the amplifier r that multiplies the activation of the go unit or the inhibition from the stop unit, depending on whether the target or fixation matches the short-term memory content. If the target matches the short-term memory content, the drift rate is $v = a \cdot r – i$; if the fixation matches the short-term memory content, it is $v = a – i \cdot r$; and if both the target and the fixation match the short-term memory content, it is $v = a \cdot r – i \cdot r = (a – i)r$. Table 1 summarizes the drift rates in the respective conditions.

In Experiments 1 and 2, we included an additional parameter (a_S) that represented the activation of the go unit in the single-task condition. Although this was an ad-hoc parameter, it appeared necessary because RT was generally shorter in the single-task condition than in the dual-task condition (see the results of Experiments 1 and 2 above). These outcomes may reflect the additional cognitive load to maintain a memory item in short-term memory in the dual-task condition. It is also possible that responding was generally slower because of switching between the prosaccade task and the memory task in the dual-task condition, yielding a switch cost that slowed responding in general. In either case, our model does not have an explicit component to explain the influence of the task load or switching. Thus, we simply assumed that the activation rate was reduced in the dual-task setting as compared to the single-task condition, with everything else remaining the same between the dual- and single-task conditions.

Fitting Procedure and Results

The model fitting procedure involved maximizing the likelihood,

$$L(\theta) = \prod f(D|\theta) \qquad (4)$$

where the likelihood function f came from Equation 2 with a set of free parameters θ and the experimental data D, which was RT for all trials in a given experiment. The model was fit by using a two-step procedure that combined a genetic algorithm and a Nelder-Mead simplex algorithm using the Matlab Optimization Toolbox (see Logan et al., 2015; Yamaguchi & Proctor, 2012). The fitting procedure started with five cycles of the genetic algorithm with the population size of 10 in each. The four best sets of the parameters at each cycle were fed back to the next cycle as part of the starting population. After the genetic algorithm, the best parameter set was used as the starting point of the simplex algorithm to find the exact location of the global maxima. All parameters were bounded at zero from below, and no parameters were bounded from above, except for t_0 whose upper bound was set at the minimum RT for the data set to which the model was fit (trials were included if RT was greater than 80 ms and less than 800 ms, to be consistent with the analyses above). The model was fit to individual participants separately as well as to all participants together as a group.

We first fit the model to the RT data of Experiments 1-3. The best parameters for individual and group fits are summarized in Table 2 (the parameters for individual fits are means across participants). The predicted and observed RT distributions are shown in Figures 6-8, respectively. The observed RT distributions for individual fits were determined based on mean parameter values. The comparisons of the predicted and observed distributions suggest the individual fits were shifted slightly to the right, as compared to the group fits, but both produced excellent fits to the data. Mean RTs were computed for the predicted and observed distributions, and the coefficient of determination (R^2) and the sum of squared errors (*SSE*) were computed for the individual and group fits (see Table 2). These mean RTs are compared in scatterplots in Figure 9. The diagonal lines in the plots represent the perfect fit between the predicted and observed mean RTs; data points above the line represent overestimations and those below represent underestimations. The individual and group fits produced similar mean RTs. As can be seen in Table 2, there were large coefficients of determination for all experiments, indicating that more than 93% of the variance in the observed mean RT was accounted for by the model predictions. There were no major discrepancies between the individual and group fits. In fact, the data points were all lined up near the diagonal lines in the respective plots. These observations demonstrate that the model predictions were impressively accurate, and they also suggest that the fitting method was not consequential.

The main purpose of the present study was to examine the influence of the short-term memory template on shifting and disengagement of spatial attention. The parameter r represented the amplifying effect of top-down process on the activation of the target and the inhibition from the fixation in a multiplicative fashion. Thus, if r is equal to one, then there is no influence of short-term memory. If *r* is greater than one, then short-term memory amplifies the activation and the inhibition at the same time. If *r* is less than one, then short-term memory depresses the activation and the inhibition; this might occur if participants have used a memory template to reject a matching object (e.g., Arita et al., 2012). As discussed in the Results section of the experiments, we have obtained significant facilitating effects of short-term memory on shifting in Experiment 1. The parameter *r* was similar between the individual fit (1.069) and group fit (1.073), and both resulted in the value of *r* that is larger than one. To confirm this, *r* for the individual fit was submitted to a one-sample *t*-test, which showed that *r* was significantly larger than one, $t(23) = 3.77$, $p = .001$, $BF_{10} = 51.12$[2]. In Experiment 2, we obtained no effect on disengagement. Consistent with this, the group fit showed *r* smaller than one (.969), but *r* for the individual fit (1.149) was numerically larger than that obtained in Experiment 1. The result of *t*-test on *r* showed that the value was significantly different from one, $t(22) = 6.83$, $p < .001$, $BF_{10} = 24039.67$, so the individual fits are inconsistent with the analysis of the experimental data. It appears that the variance was particularly larger for Experiment 2 (*SE* = .168) than for Experiment 1 (*SE* = 018), indicating large individual differences in the former. In fact, there were three participants who showed large values of *r* (> 2), and when these participants were removed, *r* turned out to be significantly smaller than one (.880), $t(19) = 11.89$, $p < .001$, $BF_{10} = 32372946$, indicating that the short-term memory template was used to reject a memory matching object. Therefore, the modeling results revealed differences in strategic use of the memory template in Experiment 2. Furthermore, in Experiment 3, we have found that the short-term memory content affected shifting toward the target and disengagement from the fixation. These observations were also supported by the model fitting outcomes. Both the individual and

---

[2] Bayes factors (BF_{10}) represents the degree to which the data support the alternative hypothesis as compared to the null hypothesis. BF was computed by using the online calculator (http://pcl.missouri.edu/bayesfactor).

group fits resulted in the same value of r (1.099), and the individual fit showed that it was significantly different from one, $t(23) = 4.68$, $p < .001$, $BF_{10} = 363.973$.

Fitting to the data of Experiment 4 was essentially the same as that for Experiment 3. Experiment 4 involved two separate phases in which the proportion of target match trials was varied. Two groups of participants also differed in the order of the two conditions. Although the results were similar between the two groups, there were some idiosyncrasies. Thus, the template model was fit separately to the two conditions of the two groups. The best parameters are summarized in Table 3, and the observed and predicted RT distributions are shown in Figure 10. Figure 11 shows scatterplots of observed and predicted mean RTs. The results demonstrated excellent fits of the template model. Note that for both groups of participants, the mostly target mismatch condition yielded little variability in mean RTs (see Figures 3b and 3d), and these experimental results were reflected in the amplifier parameter (r) that is not significantly different from one, ts < 1, $p > .4$, $BF_{10} < .2$ (see Table 3). The inhibitory parameter (i) for this condition was also very small (.005 and .007), although it was still significantly different from zero, ts > 2.7, ps < .02, $BF_{10} > 3.9$. Therefore, the model predicted identical mean RTs, which resulted in undefined R^2 and the scatterplots were cluttered at small regions (see Figure 10b and 10d). In contrast, the mostly target match condition showed results similar to Experiment 3. The amplifier (r) was significantly different from one, ts > 2.6, ps < .015, $BF_{10} > 3.4$, and R^2 was substantially large, indicating that the model predictions accounted for more than 93% of the variance in observed mean RT (see Table 3). The data points in the scatterplots also lie very close to the diagonal lines representing the perfect fit (see Figures 11a and 11c).

Overall, the modeling results support our earlier conclusion that short-term memory content can be used strategically to guide spatial attention, and when it is used, the short-term memory content affects both shifting and disengagement of attention. The model parameters are sensitive enough to discriminate the presence or absence of the top-down influence on spatial attention, and the individual fits of Experiment 2 suggest that they may be more sensitive in detecting individual differences.

Testing Theories of Attention Capture

We have so far used the template model to derive the best model parameters that could account for the experimental data most optimally. One can consider this to be a *measurement approach* as the assumptions of the model are minimal and the contributions of hypothetical cognitive factors (e.g., the top-down control in the present case) can be examined by observing the resulting model parameter values. This has allowed us to conclude that short-term memory contributed to shifting and/or disengagement of attention in Experiments 1, 3, and 4, but not in Experiment 2 (although the modeling results suggest that there were individual differences). Another approach to use computational models is to test competing theories more directly (see Yamaguchi & Proctor, 2012). To do so, the modeling framework needs to be general enough to instantiate different theories in its language. This requires deriving explicit assumptions from each theory and these assumptions to be translated into the modeming language that the framework offers. Below we take this approach to compare the two theories of attention capture, the purely stimulus-driven account and the contingent attention account.

The purely stimulus-driven account of attention capture is straightforward. It proposes that a target-defining feature does not affect shifting of attention but it slows disengagement. Thus, the match between the target and a memory template would not facilitate saccade response, whereas the match between the fixation and a memory template would slow saccade response. The amplifier r does not affect the activation a but does affect the inhibition i. The contingent attention account is also clear as to the influence of attentional set on shifting of attention. However, the account is not explicit as to whether the same attentional set also affects disengagement. We assume that attentional set only facilitates shifting but does not affect disengagement. Thus, the match between the target and a memory template would facilitate saccade response, whereas the match between the fixation and a memory template would not affect saccade response. The amplifier *r* affects the activation *a* but does not affect the inhibition *i*. The drift rates of the two accounts are summarized in Table 4.

The Wald diffusion models of the contingent attention account and the purely stimulus-driven account were fitted to the data from Experiment 3, which included both manipulations of the target and fixation match with the memory template. The outcomes of the mostly target match condition of Experiment 4 also agreed with the results of Experiment 3, but there were complex patterns across groups with different condition orders, so Experiment 3 appears to provide a better testbed. In addition to the contingent attention and purely stimulus-driven accounts, we also included the template model developed in the preceding section as a reference. The three models were compared in terms of the Bayesian Information Criterion (BIC), which is defined by

$$BIC = -2 \cdot \log[L(\theta)] + k \cdot \log[N], \quad (5)$$

where L comes from the likelihood function of Equation 4. It penalizes the model with a larger number of free parameters (*k*) according to the size of the data (*N*), but the number of parameters were the same for all three models. As BIC is proportional to the prediction error, a model with smaller BIC indicates a better fit. A difference in BIC greater than 6 is considered to be substantial evidence for a better fit; a difference greater than 10 is considered to be very strong evidence. BICs for the template model and the contingent attention model were the same (*BIC* = 53041), but they were larger for the purely stimulus-driven account (*BIC* = 53067) with the difference of 26 (see Table 5), which indicated strong evidence that the template model and the contingent attention account fit to the data better than the purely stimulus-driven account. Figure 12 also shows the comparisons of the observed RT and the predicted RT for the three models. The predictions closely overlap between the template model and the contingent attention account, and their predictions lie very closely to the diagonal line indicating a perfect fit. The predictions of the stimulus driven account deviated from the diagonal line, compared to the other two models. These results imply that the facilitation of shifting is important in predicting the experimental data. The similar performances of the template model and the contingent attention account suggest that the top-down influence on disengagement is not as strong as that on shifting. This may be due to the fact that the parameter value reflecting the inhibition from the fixation is smaller than that reflecting the activation of the go unit in the first place. We thought that it was possible that the amplification of activation and inhibition in the template model required separate parameters. We have tested this possibility, but the model fit did not improve much with the additional parameter ($BIC = 53050$) with the best-fit amplifying parameters r for activation being 1.101 and that for inhibition being 1.055. This rules out the possibility that the memory template affected activation and inhibition differently. On the other hand, we also note here that the observed top-down influence on disengagement depended on one of the eight conditions (i.e., overlap trials for fixation match) in Experiment 3. On gap trials, fixation match had little influence because the temporal gap between fixation offset and target onset allowed pre-disengagement. Although the effect was detected in Experiment 3 (and in Experiment 4 as well), these subtle differences between the two models did not outstand when summary statistics like BIC are compared.

To further distinguish between the contingent attention account and the template model, we performed the Bayesian model selection for group data (Stephan, Penny, Daunizeau, Moran, & Friston, 2009)[3]. This procedure started with a 5-fold cross-validation,

---

[3] We thank Yu-Chin Chiu for suggesting this procedure.

in which each participant's data were partitioned into the training and validation datasets, and a model was fitted to the training dataset to derive the best parameter set. This parameter set was then used to obtain the goodness-of-fit measure (i.e., BIC) of the model to the validation dataset. For each participant, we sampled 80% of the data from Experiment 3 randomly as the training dataset, and the remaining 20% were used as the validation dataset. This procedure was repeated five times for each participant, and the goodness-of-fit measures for five iterations were averaged. This cross-validation resulted in 24 mean goodness-of-fit scores (representing 24 participants) for each of the two models. Finally, the exceedance probability was computed for the two models[4] (Stephan et al., 2009). This probability represents the likelihood of a model given the group data against all other models that were compared. In our case, the exceedance probability of the contingent attention account was $p = .995$, and that of the template model was $p = .005$. That is, the likelihood of the contingent attention account against the template model was 95.5% (and that of the template model was .5%). The results provided overwhelming support for the contingent attention account, suggesting that top-down influences on disengagement did not contribute to the overall goodness of fit to the present data set. Implications of these results are discussed in the General Discussion.

General Discussion

The mechanism underlying the top-down guidance of attention capture is still a subject of active debate. The contingent attention account suggests that top-down processes establish an attentional set to filter visual information that is irrelevant to detecting the target object, and attention capture occurs only if the exogenous cue contains the target-defining feature (Folk et al., 1992). This account would predict that top-down processes affect the shifting component of attention orienting. On the other hand, the purely stimulus-driven

---

[4] https://github.com/canlab/spm8/blob/master/spm_BMS.m

account suggests that attention capture occurs whenever a perceptually salient object is presented, regardless of whether it contains the target-defining feature, and it refutes the possibility that there is a role of top-down processes in attention shifting (Theeuwes, 2010). Instead, top-down processes can delay disengagement from a peripheral cue if the object contains the target-defining feature. Furthermore, the dwelling account agrees with the purely stimulus-driven account that attention capture occurs in a stimulus-driven manner, but disengagement from distractors (not the peripheral cue) can be slowed when search is difficult (Gaspline et al., 2016). This account does seem to suggest that top-down processes affect disengagement but is mute as to whether it also affects shifting. These accounts provide three different views of the top-down influence on attention capture, but none of them seem to account for the results of the four experiments of the present study fully.

In Experiment 1, participants looked to the target that appeared on the left or right of the fixation. The target occurred abruptly and should have captured attention in a stimulus-driven manner. Responses were faster when the target color matched the content of short-term memory than when it did not. This result implies that top-down processes accelerated the stimulus-driven attention capture (Reinhart et al., 2016), demonstrating the top-down contribution to attention shifting. Although Experiment 2 did not provide evidence that top-down processes affected attention disengagement, Experiment 3 showed that the gap effect was larger when the fixation color matched the short-term memory content than when it did not match. This outcome indicated that it took longer to disengage from the fixation in the former condition than in the latter condition. Therefore, short-term memory does impair attention disengagement when participants decide to use the short-term memory content to guide attention toward the target. Experiment 4 further confirmed strategic use of short-term memory template by varying its utility (i.e., the proportion of trials for which the short-term memory content matched the target). Top-down influences on shifting and disengagement were obtained when the utility of short-term memory was high, but not when it was low. When the utility of short-term memory for accelerating attention shifting was low in this experiment, participants even appeared to have ignored the fixation that could impair disengagement when its color matched with the short-term memory template. Although this outcome could still reflect a strategy adopted by participants, it was a different type of strategy that was beyond the scope of the accounts of top-down influences we considered in the present study. Therefore, further investigations are required to clarify different types of strategies that could be utilized in the control of visual attention.

Our implementations of the template model fit well to the experimental data from the four experiments. The model assumed a race between the go and stop units that represented the activation and inhibition of prosaccade responses, respectively (Boucher et al., 2007; Logan et al., 2015). The top-down influence was represented by an amplification of the strengths of activation and inhibition, and the model successfully differentiated the presence and the absence of the top-down guidance in the four experiments. Namely, the model indicated that the top-down guidance was present in Experiments 1, 3, and 4 of the mostly target match condition, but not in Experiments 2 and 4 of the mostly target mismatch condition; the model also suggested that there were individual differences in the use of short-term memory in Experiment 2. The modeling approach was also used to test the contingent attention account and the purely stimulus-driven account with the data from Experiment 3. The contingent attention account produced a better fit to the model than the purely stimulus-driven account, which indicated an important role of amplification of the go unit activation when the target color matched the memory template. Interestingly, an amplification of the inhibition did not improve the overall model fit, at least, in terms of the BIC measure used in the present study. The follow-up Bayesian model selection for group data provided overwhelming support for the contingent attention account over the template model.

Therefore, computational modeling indicated that top-down guidance was effective mainly for the activation of saccade responses but may not so for the inhibition of these responses.

Nevertheless, these modeling results require some caution to interpret. Model selection was performed in terms of the goodness of fit to the overall data (i.e., BIC). Careful considerations of the experimental data reveal that the top-down influence on disengagement was reflected only in one of the data points (overlap trials for fixation match) in Experiments 3 and 4. Fortunately, the effect was detected in terms of the statistical inferences on the experimental data, but it might have been concealed when computational models were compared in terms of summary statistics, such as BIC. Therefore, further scrutiny is needed to examine the top-down contributions to disengagement in an experimental design that reveal a greater contribution of disengagement to the overall task performance. Such a study is currently on-going in our lab, using a procedure in which successful task performance requires strong contributions of inhibitory control on saccadic reactions. The results of this study may provide unambiguous top-down contributions to attention disengagement.

There are a few directions that the present modeling approach can be extended. First, the present model can be implemented within the LCA formalization as in the original interactive race model (Boucher et al., 2007; Logan et al. 2015). This requires a high performance computing system, which is not always available to everyone. Furthermore, the present model focused on response activation, and future work may interface the activation component with a formal model of encoding (see Figure 4a). Such a development is most relevant when rich, complex visual information is available to the model. There are many computational models that have focused on the encoding part of the theory, but these models do not predict details of RT distributions as in the present approach (Borji & Itti, 2013). Thus, integrating these approaches can potentially be useful to account for performance with complex visual displays. Furthermore, other types of gaze behaviors could be incorporated into the model, such as gaze trajectories and fixation. These developments would make the current model more applicable to the world outside the laboratory.

Memory Template or Dimensional Weighting?

Does the present study end the debate as to which components of attention orienting depend on top-down processes? We believe that the present study has settled the debate within the paradigm that relies on a short-term memory template to guide visual attention (Downing, 2000). Nevertheless, one can still argue that there are different mechanisms by which top-down processes affect attention capture according to the task setting. For instance, researchers distinguish between early and late attentional selection processes (e.g., Lavie, Hirst, de Fockert, & Viding, 2004; Treisman, 1969). The early selection depends on a perceptual filter that blocks visual input, and the late selection resolves interference from irrelevant visual stimuli depending on the availability of cognitive resources. The contingent attention account would suggest that a top-down process sets up an attentional filter at early selection (Folk et al., 1992). In contrast, the technique used in the present experiment relied explicitly on short-term memory, which can be thought to be a part of the late selection mechanism (Lavie et al., 2004). Whether the content of short-term memory directly affects the attentional filter or only reduces interference at the later selection process is still subject to further investigations. There may be different loci of top-down influences on stimulus-driven attention capture. A future challenge is to dissociate top-down influences that may arise from different mechanisms.

Conclusion

The present study contributed to the current debate as to the nature of top-down influences on stimulus-driven attention capture. The experimental results demonstrated that top-down processes can accelerate attention shifting by using a memory template in short-term memory, but there was not much support for a direct influence of top-down processes on disengagement operations if there is no incentive to use the short-term memory content to guide attention toward the target. When the use of the short-term memory template can facilitate shifting to the target, top-down processes do impair disengagement of attention from the current fixation that matches the short-term memory template. Hence, the present results are clear in showing top-down influences on attention shifting and disengagement and their reliance on strategic use of a short-term memory template. The computational implementations of the template model demonstrated that the model was able to discriminate experimental conditions in which the memory template was used to guide attention. The modeling results further suggested that the top-down contribution to shifting is more prominent than that to disengagement. The current modeling approach provides a useful step forward toward a more complete model of visual attention. Future developments should embody the model with more sophisticated encoding mechanisms and a complete oculomotor system that allows predictions of more detailed oculomotor parameters, such as saccade trajectories and fixation durations.

Context

The study of visual attention is one of the main lines of research in experimental psychology that has made an impact in a variety of sub-disciplines. The study also has a long tradition in bridging between basic and applied research, exemplified by the works of Donald Broadbent, William Hick, Alan Baddeley, and many others. It is now widely recognized that the understanding of attention control is important in characterizing psychopathological conditions, and many forms of attention tasks have been developed and are being used to assess patients in clinical contexts. Deficits of attentional control are also known to be an important issue to be addressed in educational contexts, which may be relevant to populations ranging from school children to elderlies. The present study used a new experimental paradigm and cutting-edge computational modeling techniques to address one of the contemporary issues in the study of attentional control. The results provided an insight into how two forms of attentional control interplay to enable efficient reactions to changing environments. The present findings may be extended by asking questions such as how the interactions between the two systems develop throughout the lifespan and whether different pathological conditions affect the control ability uniquely. Answers to these questions would contribute to improving well-being within general and special populations.

References

Anderson, B. A., & Folk, C. L. (2012). Dissociating location-specific inhibition and attention shifts: Evidence against the disengagement account of contingent capture. *Attention, Perception, & Psychophysics, 74*, 1183-1198.

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: Configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 580-584.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attention capture. *Perception and Psychophysics, 55*, 485-496.

Broji, A., & Itti, L. (2013). State-of-the-art in visual attention modelling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*, 185-207.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice reaction time. *Psychological Review, 112*, 117-128.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice reaction time: Liner ballistic accumulation. *Cognitive Psychology, 57*, 153-178.

Boucher, L., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2007). Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review, 114,* 376-397.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*, 523-547.

Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta Psychologica, 137*, 217-225.

Carrasco, M., (2011). Visual attention: The past 25 years. *Vision Research, 51,* 1484-1525.

Cave, K. R., & Wolfe, J. M., (1990). Modelling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225-271.

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

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28-71.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Review Neuroscience*, *3*, 201-215.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*, 306-324.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193-222.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*, 467-473.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433-458.

Egeth, H. E., & Yantis, S. (1998). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269-297.

Fischer, B., & Breitmeyer, B. (1987). Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia*, *25*, 73-83.

Folk, C. L., & Remington, R. (2006). Top-down modulation of preattentive processing: Testing the recovery account of contingent capture. *Visual Cognition*, *14*, 445-465.

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

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

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a “dimension-weighting” account. *Perception & Psychophysics*, *58*, 88-101.

Gaspelin, N., Ruthruff, E., & Lien, M.-C. (2016). The problem of latent attentional capture: Easy visual search conceals capture by task-irrelevant abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1104-1120.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert visual attention. *Vision Research*, *40*, 1489-1506.

Jin, Z., & Reeves, A. (2009). Attentional release in the saccadic gap effect. *Vision Research*, *49*, 2045-2055.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*, 1-32.

Jonides, J. (1981). Voluntary vs. automatic control over the mind’s eye’s movement. In J. B. Long, & A. D. Baddeley (Eds.), *Attention and Performance IX* (pp. 187-203). Hillsdale, NJ: Earlbaum.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Review Neuroscience*, *1*, 91-100.

Kingstone, A., & Klein, R. M. (1993). Visual offsets facilitate saccadic latency: Does predisengagement of visuospatial attention mediate this gap effect? *Journal of Experimental Psychology: Human Perception and Performance*, *19,* 1251-1265.

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

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*, 339-354.

Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review, 121,* 66-95.

Logan, G. D., Yamaguchi, M., Schall, J. D., & Palmeri, T. J. (2015). Inhibitory control in mind and brain 2.0: Blocked-input models of saccadic countermanding. *Psychological Review, 122,* 115-147. doi: 10.1037/a0038893

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*, 1183-1198.

Mackeben, M., & Nakayama, K. (1993). Express attentional shifts. *Vision Research, 33,* 85-90.

Müller, H. J., & Krummenacher, J. (2006). Locus of dimension weighting: Preattentive or postselective? *Visual Cognition, 14,* 490-513, DOI: 10.1080/13506280500194154

Müller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315-330.

Oliva, A., & Torralba, A. (2001). Modeling the spahe of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145-175.

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

Pratt, J., Bekkering, H., & Leung, M. (2000). Estimating the components of the gap effect. *Experimental Brain Research, 130,* 258-263.

Pratt, J., Lajonchere, C. M., & Abrams, R. A. (2006). Attentional modulation of the gap effect. *Vision Research, 46*, 2602-2607.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Reinhart, R. M. G., McClenahan, L. J., & Woodman, G. F. (2016). Attention's accelerator. *Psychological Science, in press.*

Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America*, *57,* 1024-1029.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*, 283-317.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*, 1004-1017.

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

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

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 (Ed.), *Attention & Performance*¸ *Vol. 18*. (pp. 105-125). Cambridge, MA: MIT Press.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*, 282–299.

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

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202-238.

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99-119). New York: Oxford.

Yamaguchi, M., & Proctor, R. W. (2012). Multidimensional vector model of stimulus-response compatibility. *Psychological Review*, *119*, 272-303.

Yantis, S. (1998). Control of visual attention. In H. Pashler (Ed.), *Attention* (pp. 223-256). East Essex, UK: Psychology Press.

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 661-676.

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.

Table 1. Drift rates for the template model; a = activation of the target; i = inhibition from the fixation; r = top-down attention amplifying the activation or inhibition (a_1 is the activation parameter for single-task).

Condition	Gap trial	Overlap trial
Single task	$v = a_1$	$v = a_1 - i$
Target mismatch, fixation mismatch	$v = a$	$v = a - i$
Target match, fixation mismatch	$v = a \cdot r$	$v = a \cdot r - i$
Target mismatch, fixation match	$v = a$	$v = a - i \cdot r$
Target match, fixation match	$v = a \cdot r$	$v = (a - i)r$

Table 2. Best parameters and goodness-of-fit measures (R^2 and *Sum of Squared Errors*, or SSE) for individual and group fits of the template model in Experiments 1-3.

		a	i	z	t_0	r	a_s	R^2	*SSE*
Experiment 1	Individual	.207	.074	17.672	67.491	1.069	.264	.933	241
	Group	.145	.055	12.738	73.188	1.073	.189	.936	226
Experiment 2	Individual	.236	.063	21.164	61.210	1.149	.270	.989	52
	Group	.148	.048	12.609	72.509	0.969	.175	.977	91
Experiment 3	Individual	.223	.052	19.626	61.987	1.099	-	.975	124
	Group	.146	.034	12.706	71.827	1.099	-	.974	69

Table 3. Best parameters and goodness-of-fit measures (R^2 and *Sum of Squared Errors*, or SSE) for individual and group fits of the template model in Experiment 4.

Group	Condition		*a*	*i*	*z*	t_0	*r*	R^2	*SSE*
Mostly target match first	Mostly target match	Individual	0.230	0.051	19.825	65.972	1.096	0.937	263
		Group	0.153	0.033	12.838	72.891	1.108	0.940	88
	Mostly target mismatch	Individual	0.195	0.005	17.145	68.724	1.018	-	110
		Group	0.142	0.001	12.454	73.498	0.999	-	178
Mostly target mismatch first	Mostly target match	Individual	0.226	0.049	16.662	68.695	1.120	0.964	94
		Group	0.152	0.036	11.626	74.944	1.136	0.964	107
	Mostly target mismatch	Individual	0.200	0.007	15.511	70.152	1.005	-	129
		Group	0.142	0.000	11.339	75.138	1.001	-	153

Table 4. Drift rates for the purely stimulus-driven account and for the contingent attention account.

Condition	Stimulus-Driven		Contingent Attention	
	Gap trial	Overlap trial	Gap trial	Overlap trial
Target match, fixation match	$v = a$	$v = a - i \cdot r$	$v = a \cdot r$	$v = a \cdot r - i$
Target mismatch, fixation match	$v = a$	$v = a - i \cdot r$	$v = a$	$v = a - i$
Target match, fixation mismatch	$v = a$	$v = a - i$	$v = a \cdot r$	$v = a \cdot r - i$
Target mismatch, fixation mismatch	$v = a$	$v = a - i$	$v = a$	$v = a - i$

Table 5. Best parameters and goodness-of-fit measures (R^2, *Sum of Squared Errors [SSE], and Bayesian Information Criterion* [BIC]) for group fits of the template model, the contingent attention account, and the purely stimulus driven account.

	a	*i*	*z*	t_0	*r*	R^2	*SSE*	*BIC*
Template	0.146	0.034	12.706	71.827	1.099	0.974	69	53041
Contingent attention	0.146	0.036	12.705	71.828	1.100	0.973	71	53041
Stimulus driven	0.152	0.035	12.642	71.924	1.053	0.809	337	53067

Figure 1. The sequence of the displays for the single- and dual-task conditions. The example shows the displays of Experiment 1. In Experiment 2, the fixation was colored in the dual-task condition. In Experiments 3 and 4, both the fixation and the target were colored.

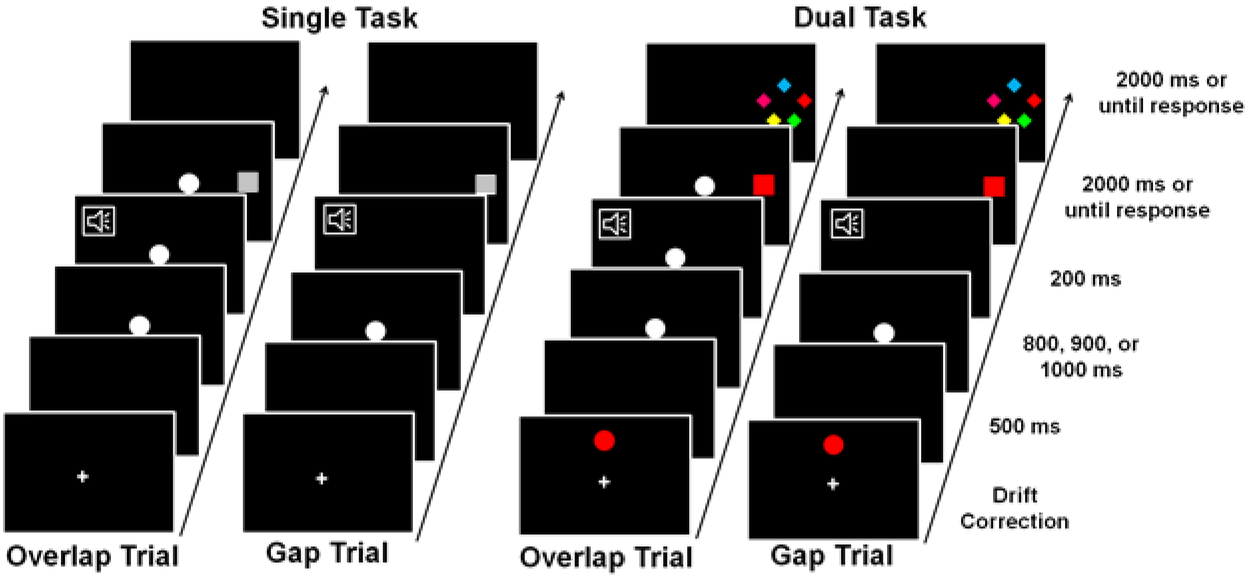
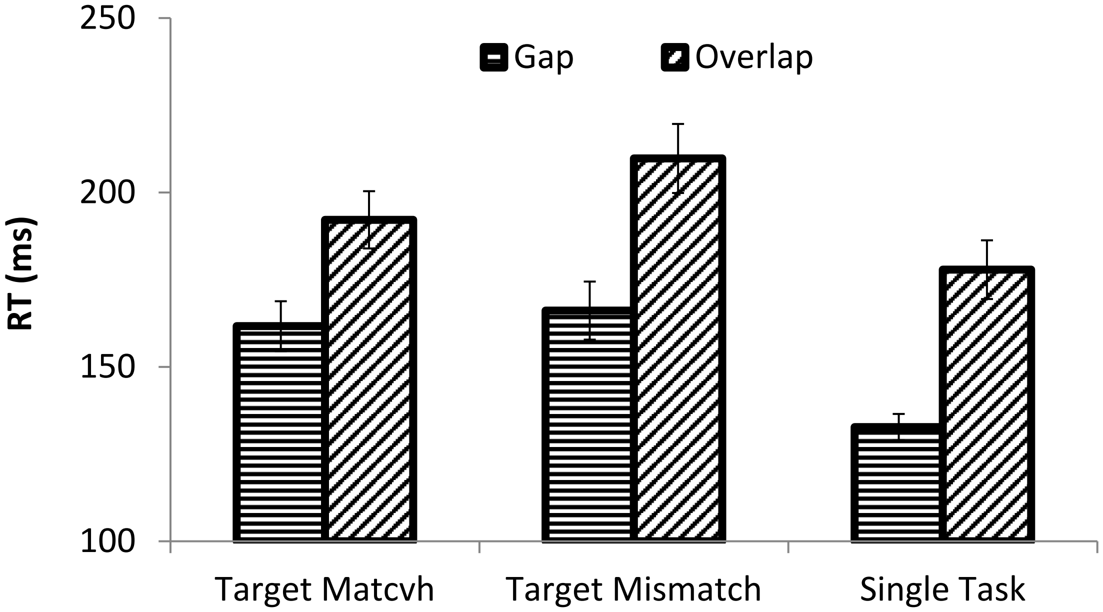
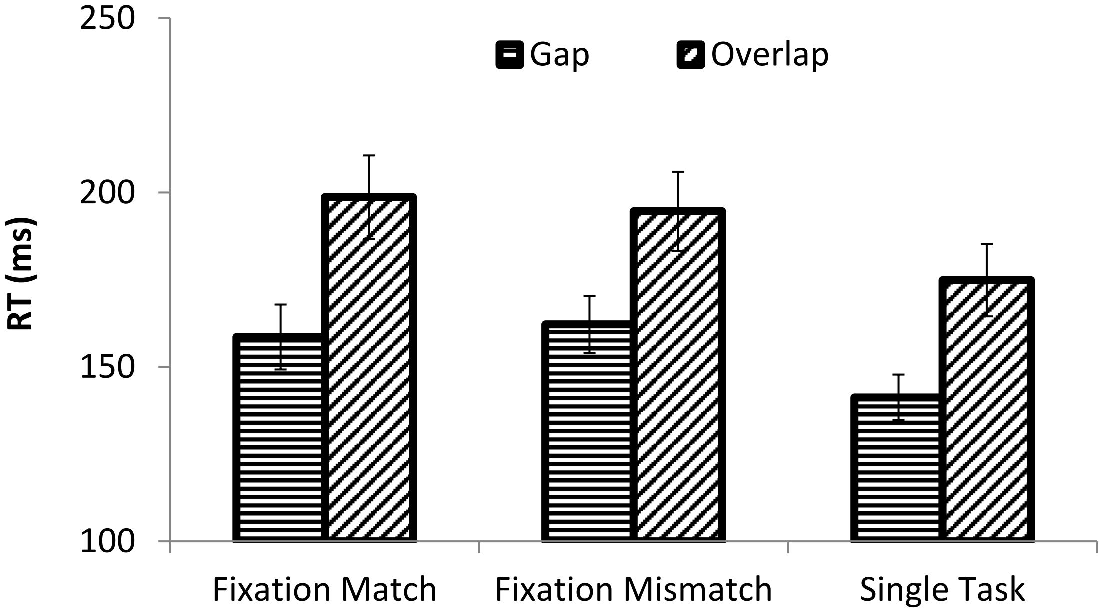


Figure 2. Mean Response Time (RT) in Experiments 1-3.

a. Experiment 1



b. Experiment 2



c. Experiment 3

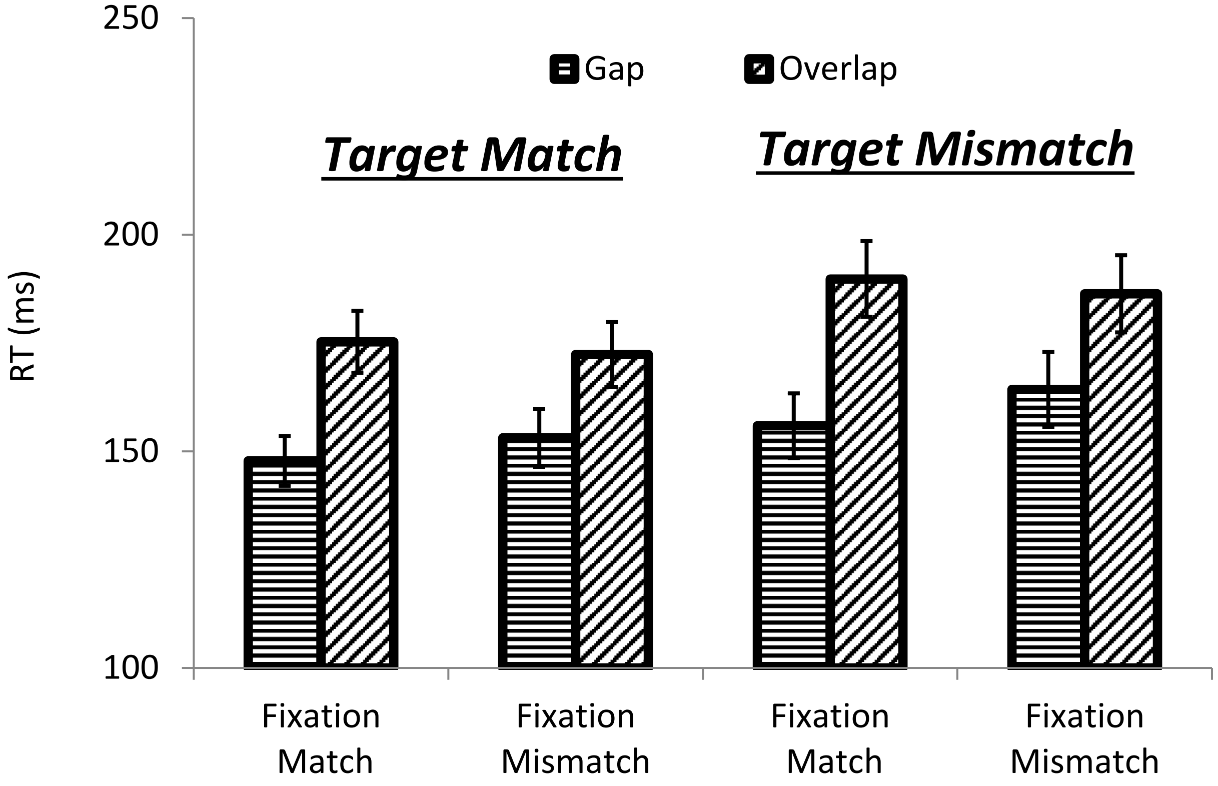
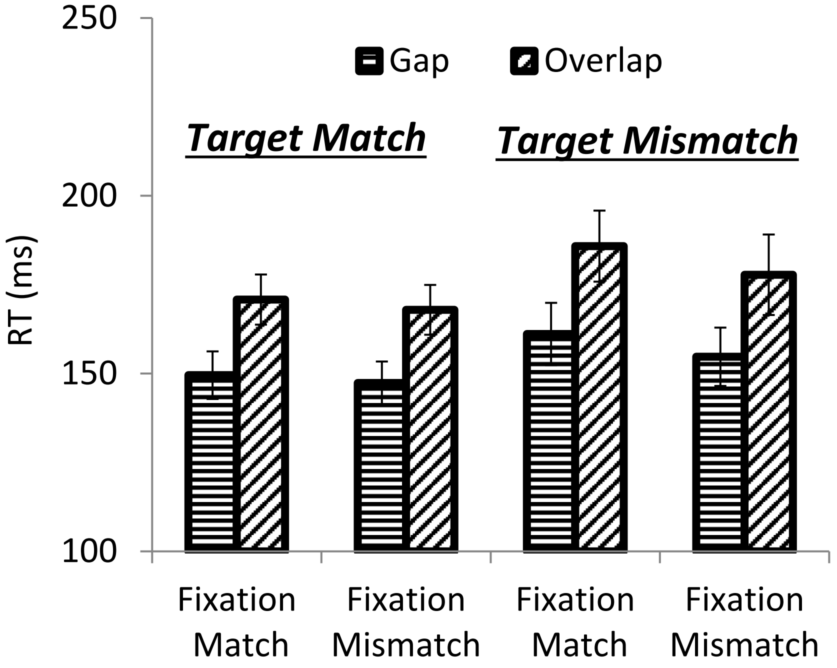
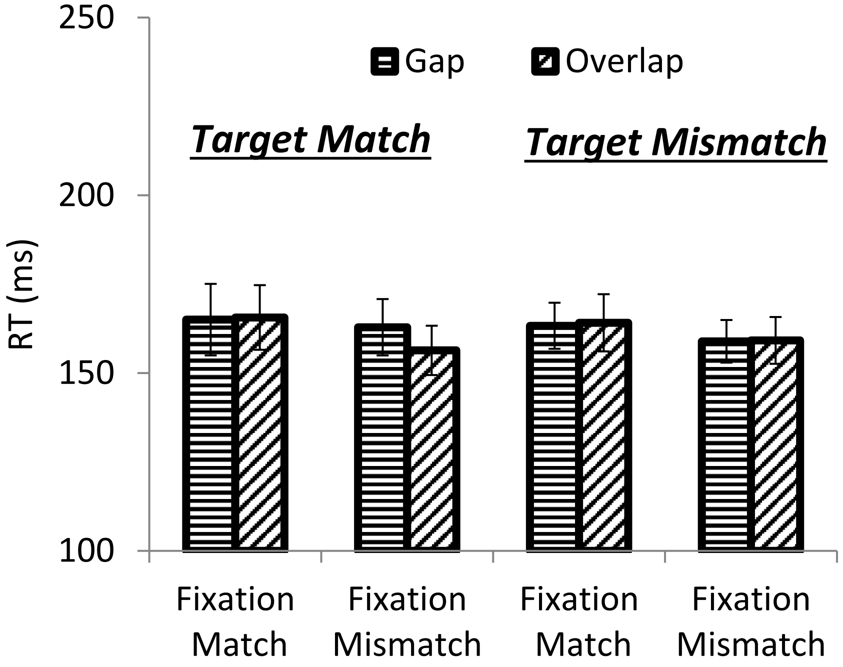


Figure 3. Mean Response Time (RT) in Experiment 4.

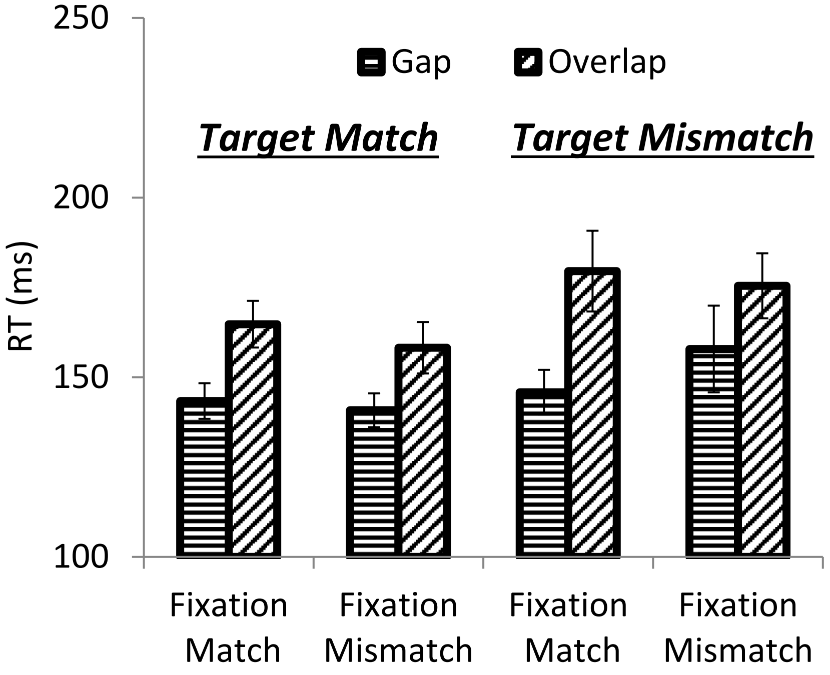
a. Group with the Mostly Target Match First: Mostly target match condition



b. Group with the Mostly Target Match First: Mostly target mismatch condition



c. Group with the Mostly Target Mismatch First: Mostly target match condition



d. Group with the Mostly Target Mismatch First: Mostly target mismatch condition

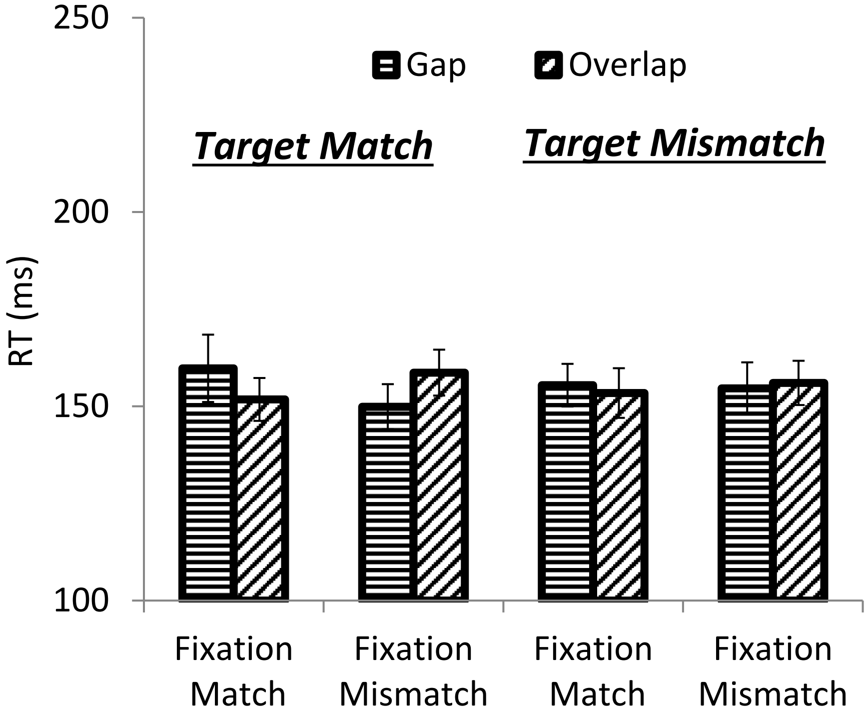
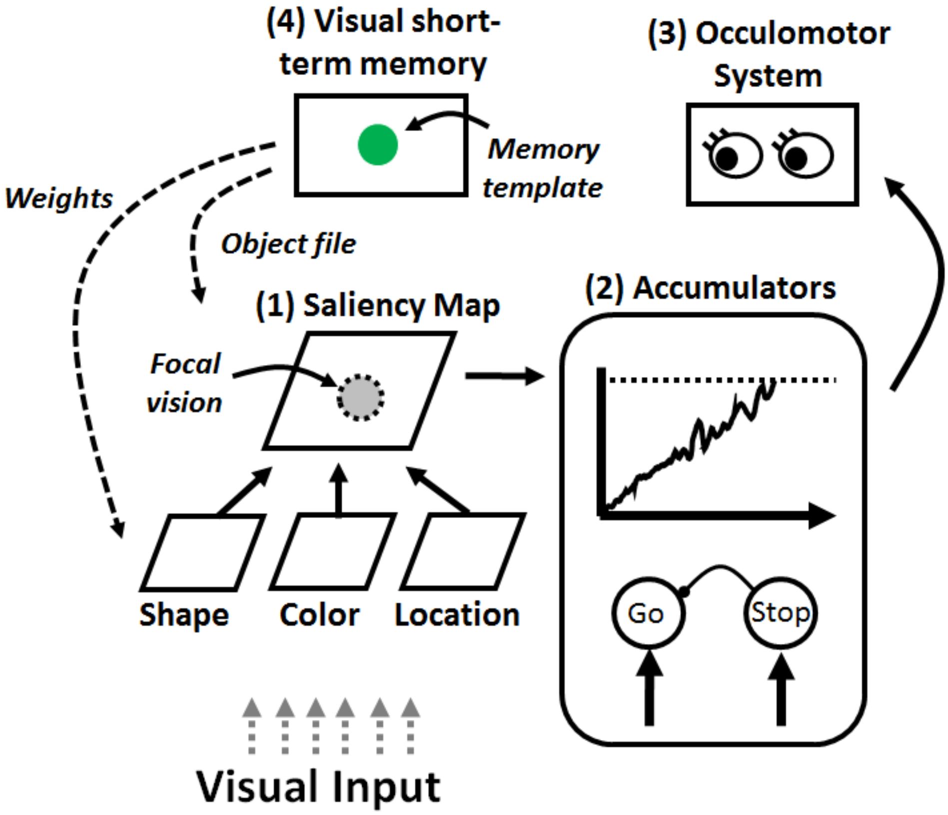
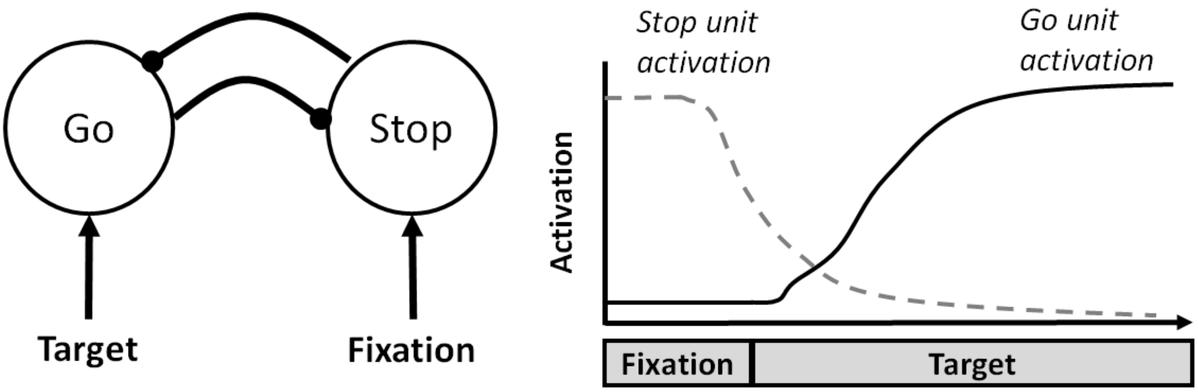


Figure 4. The model of spatial attention control. (a) The architecture of top-down control of spatial attention and saccade response. (b) The refined interactive race model by Logan et al. (2015), and the activation dynamics of the Go and Stop units during the fixation and after the target onset. (c) Wald diffusion process.

a.



b.


Go
Stop
Target
Fixation
Stop unit activation
Go unit activation
Activation
Fixation
Target

c.

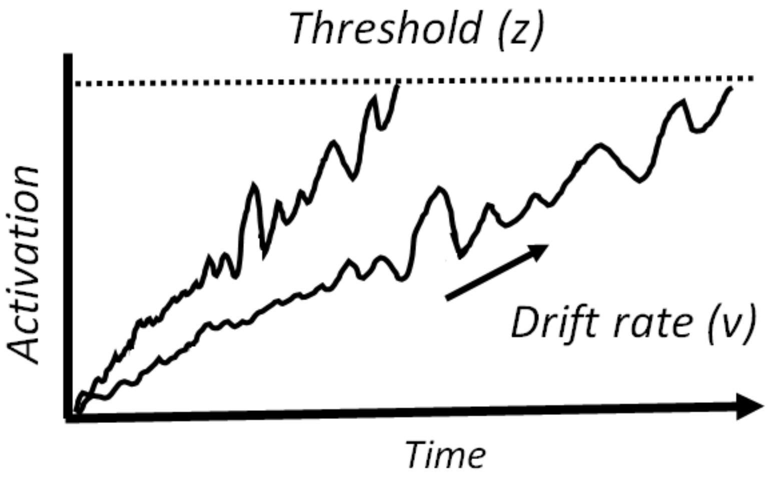
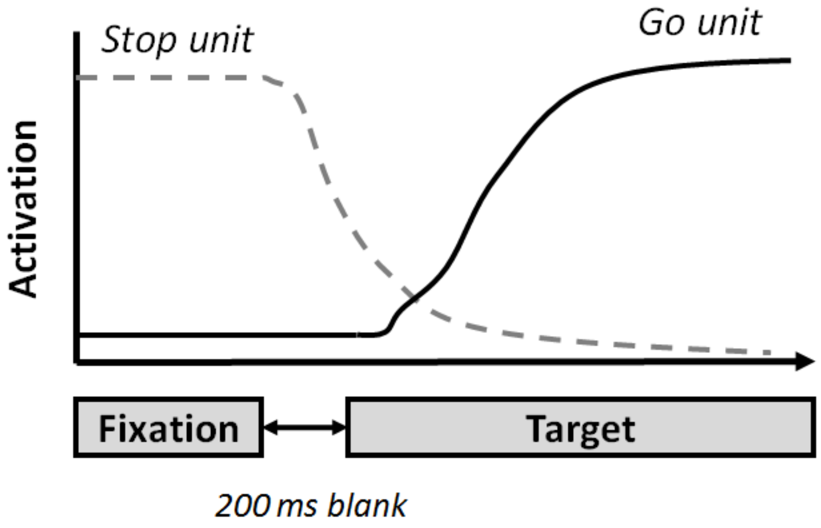

Threshold (z)
Activation
Drift rate (v)
Time

Figure 5. The interactive race model of gap and overlap trials.

a. Gap Trial



b. Overlap Trial

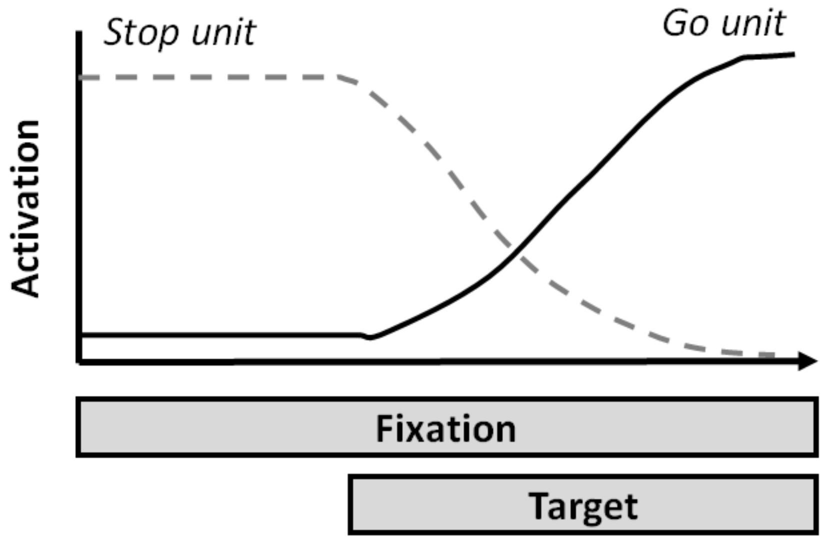


Figure 6. The predicted response time distributions for individual and group fits, and the observed data in Experiment 1. The predicted RT distributions for individual fits are dotted lines, and the predicted RT distributions for group fits are solid lines. The normalized histograms are the experimental data including all participants’ trials.

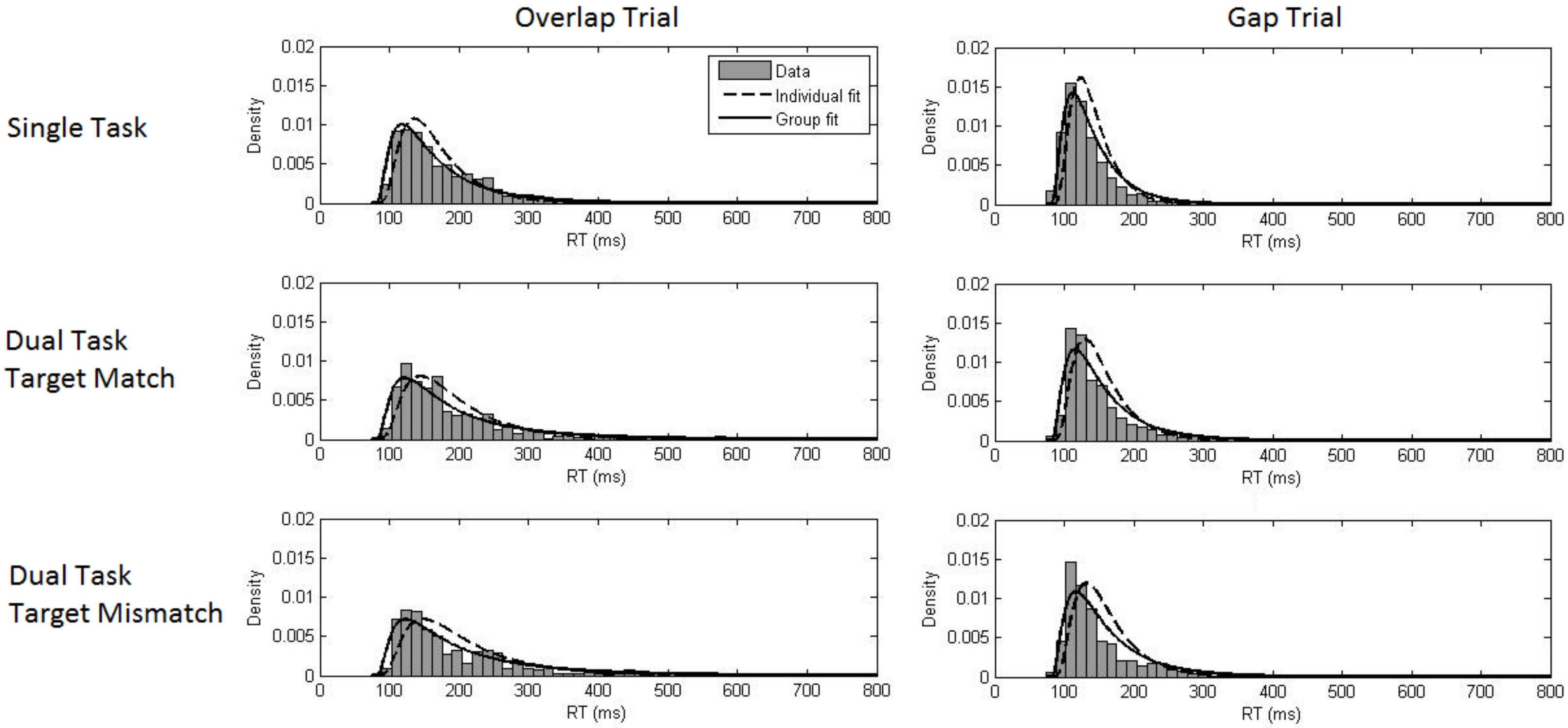


Figure 7. The predicted response time distributions for individual and group fits, and the observed data in Experiment 2. The predicted RT distributions for individual fits are dotted lines, and the predicted RT distributions for group fits are solid lines. The normalized histograms are the experimental data including all participants' trials.

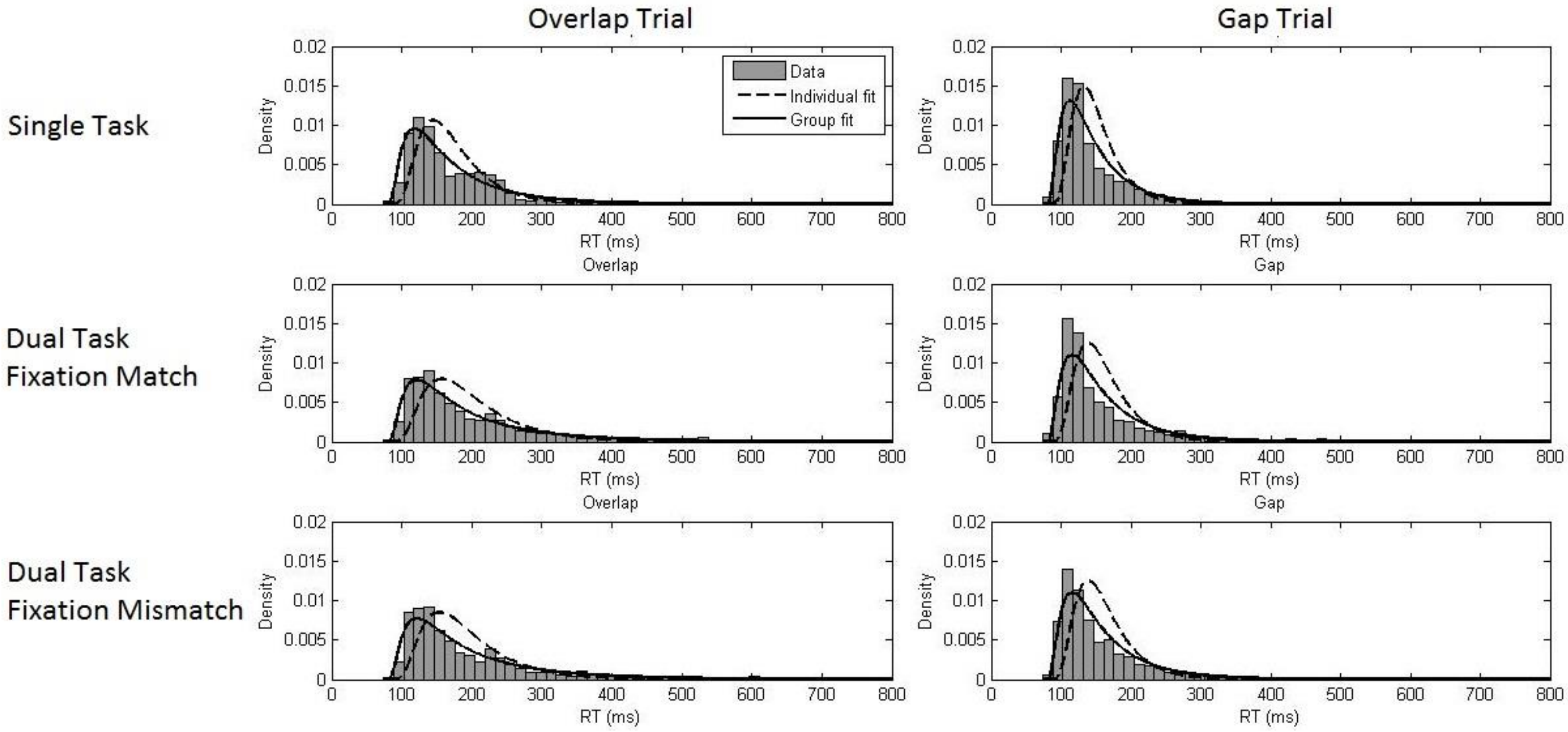
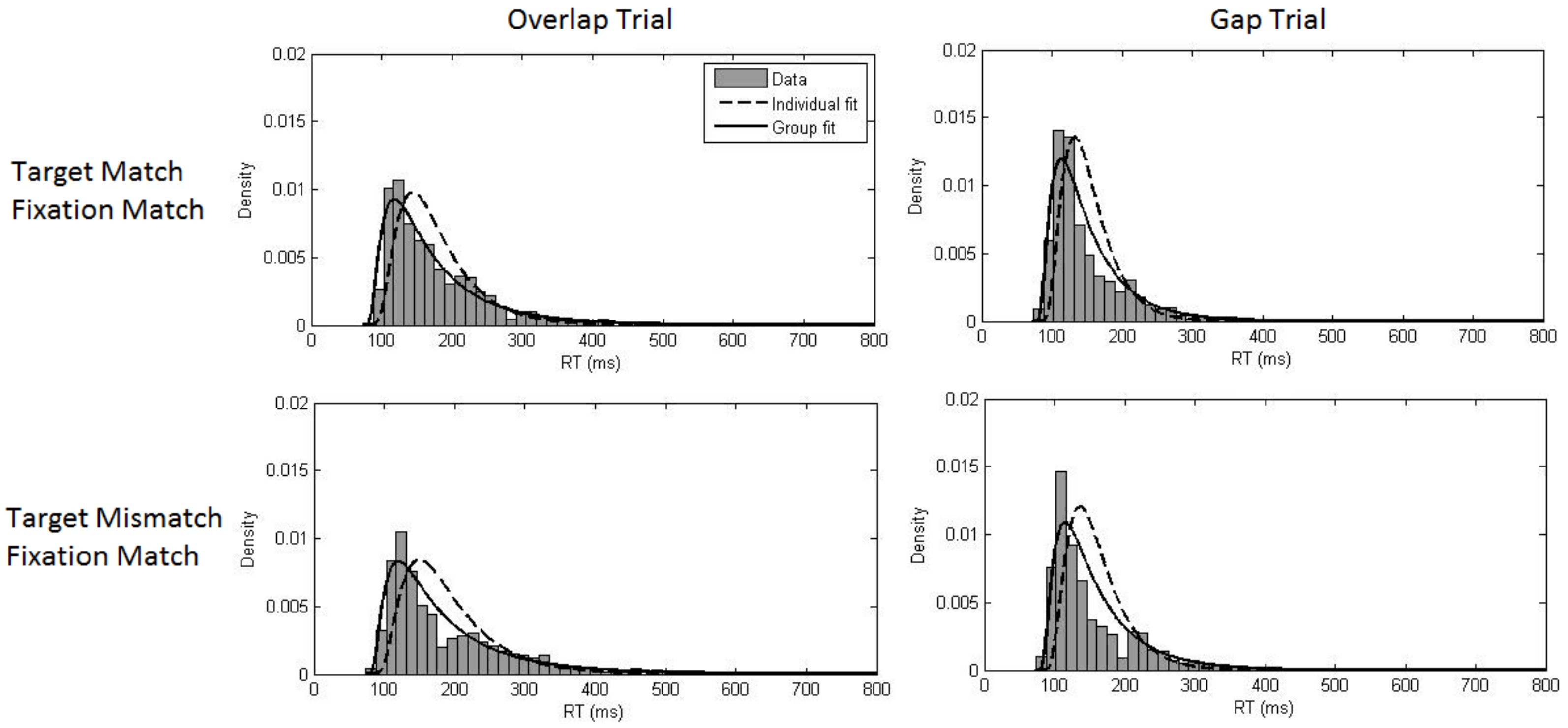


Figure 8. The predicted response time distributions for individual and group fits, and the observed data in Experiment 3. The predicted RT distributions for individual fits are dotted lines, and the predicted RT distributions for group fits are solid lines. The normalized histograms are the experimental data including all participants' trials.



(continued)

(Figure 8 continued)

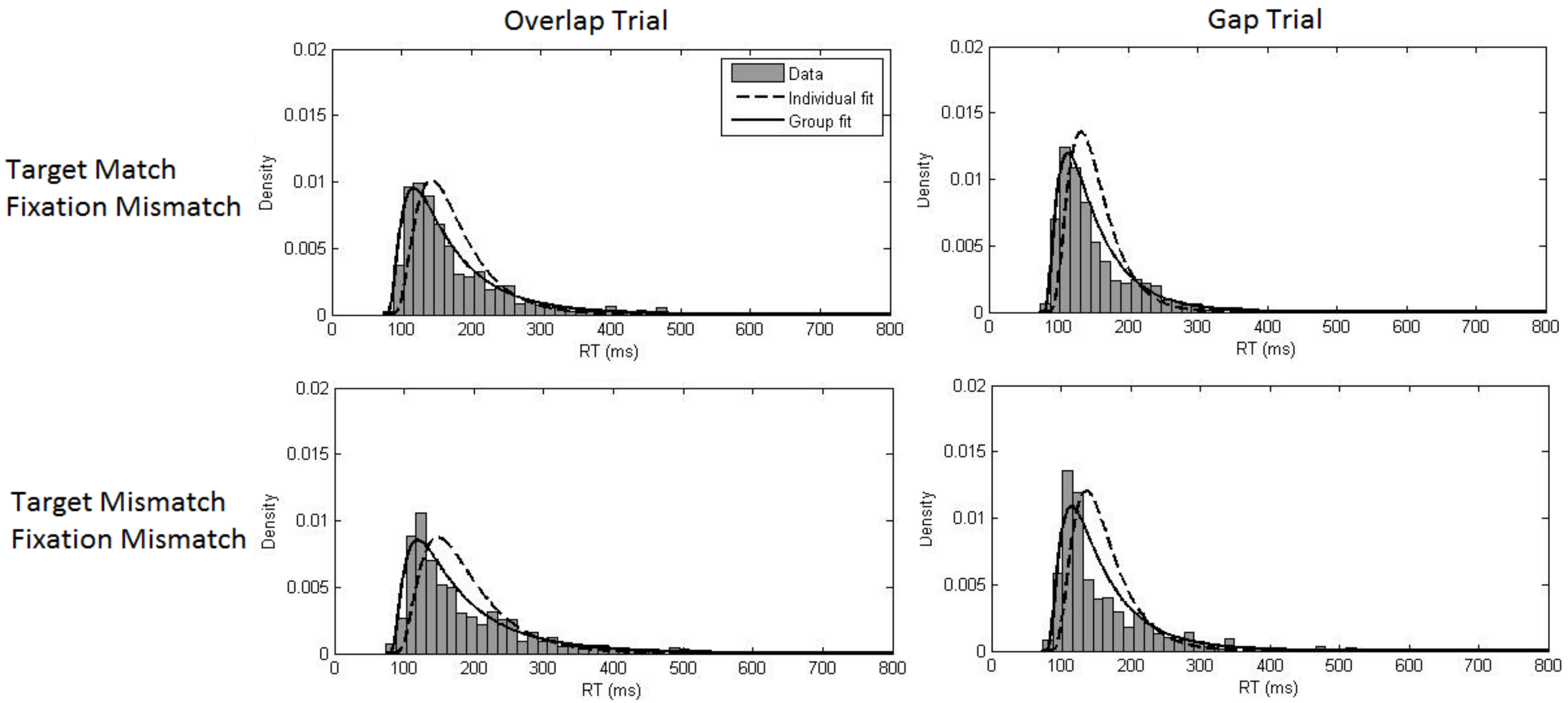
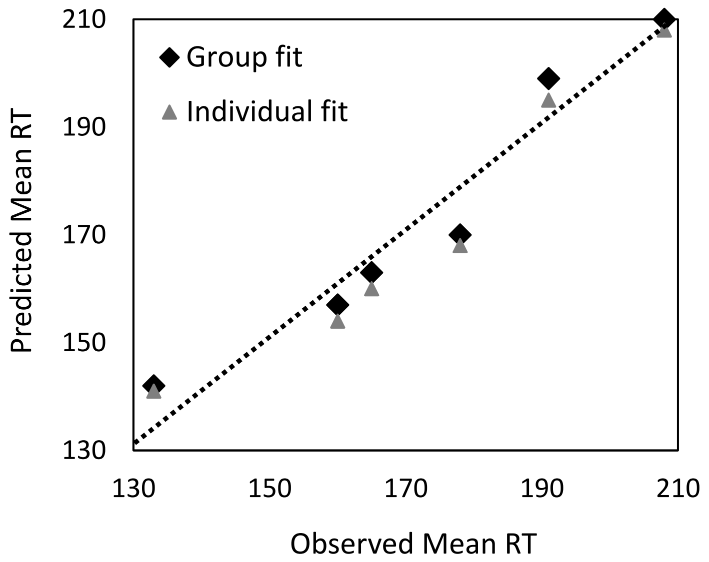
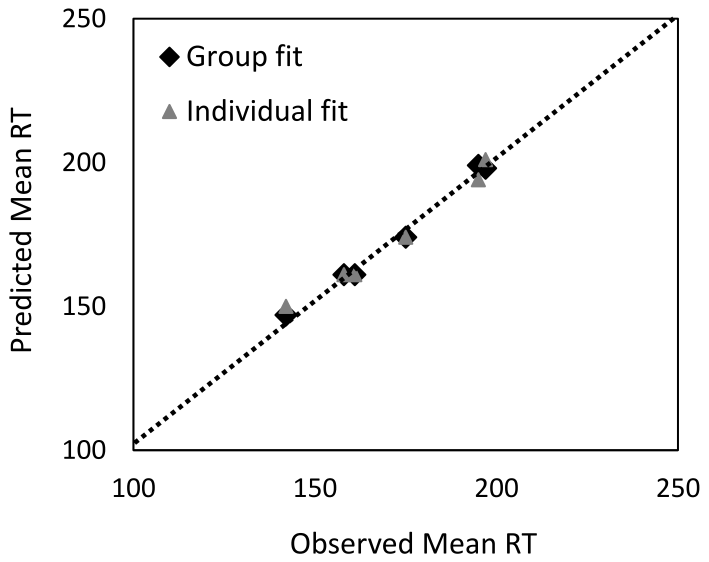


Figure9. Mean response times (RTs) for the individual and group fits in Experiments 1-3. The dotted diagonal lines represent the perfect fit.

a. Experiment 1



b. Experiment 2



c. Experiment 3

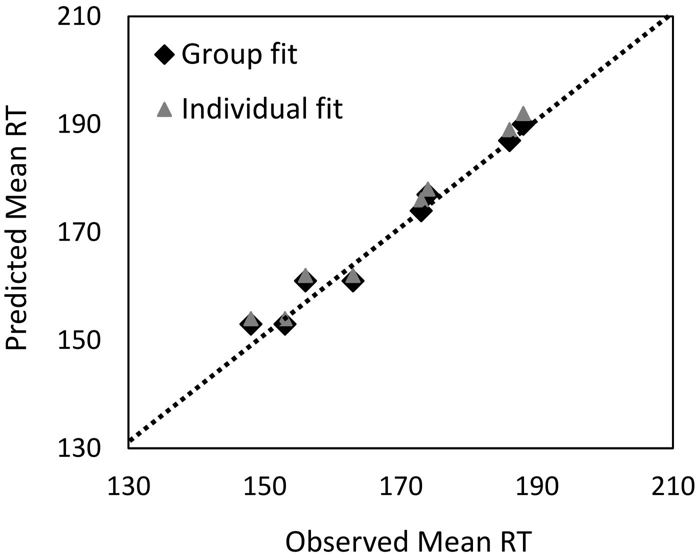
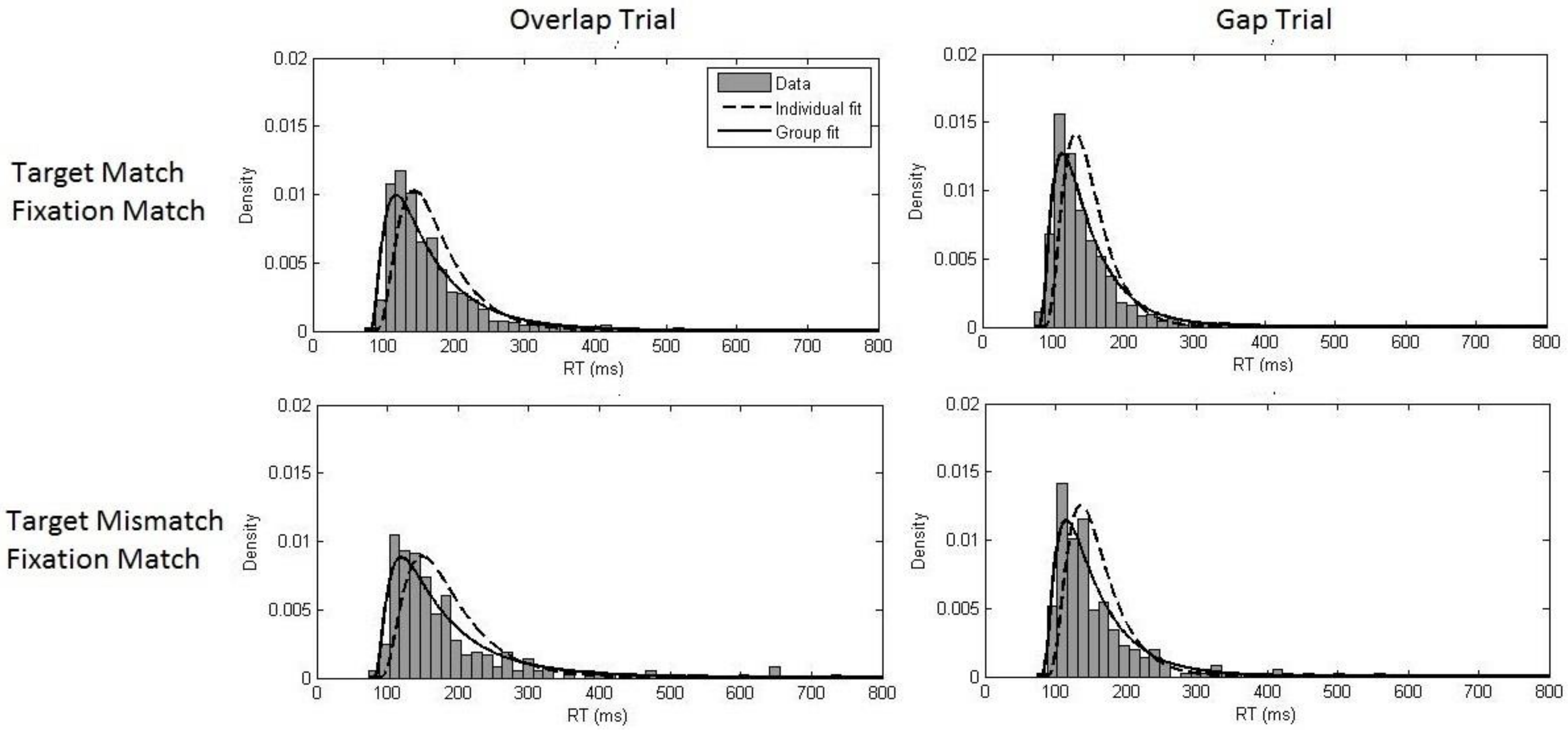
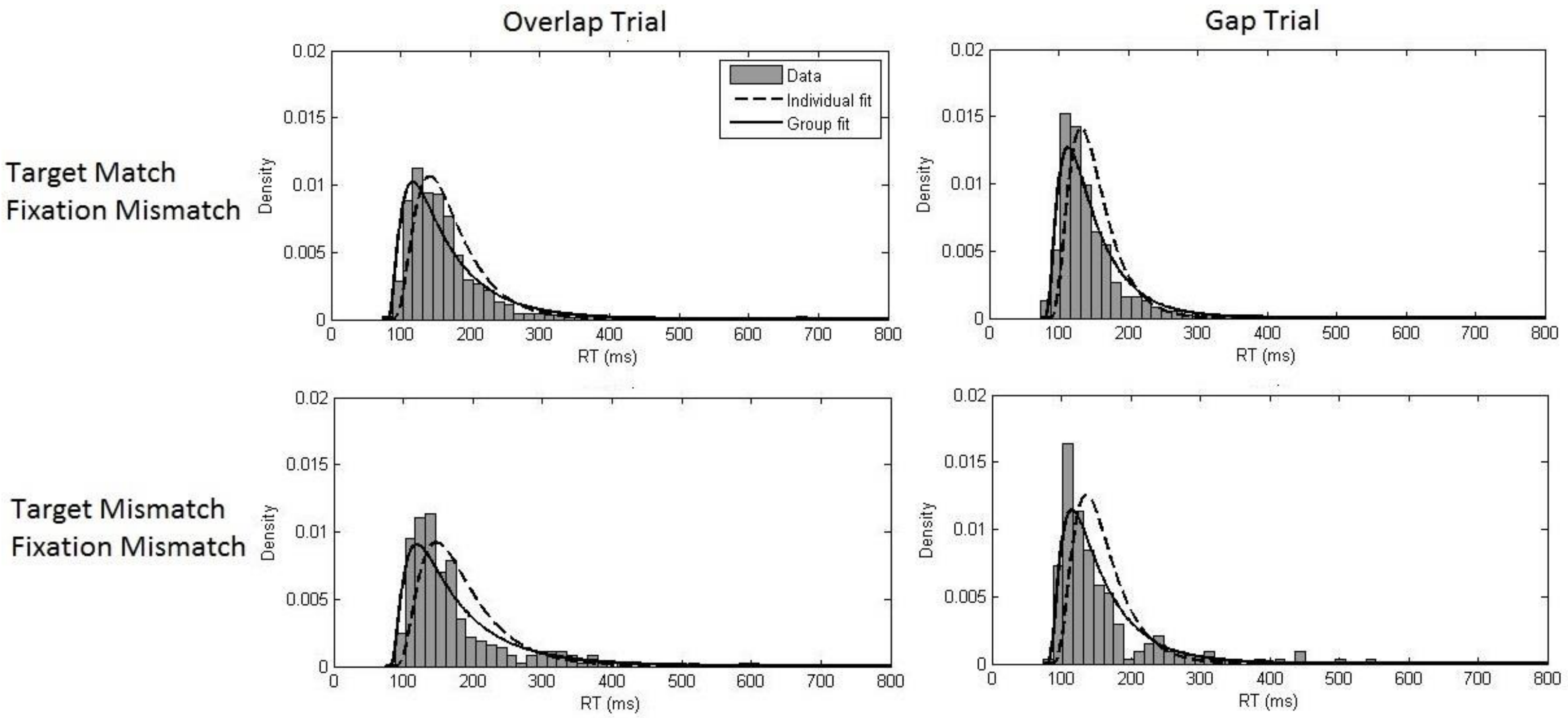


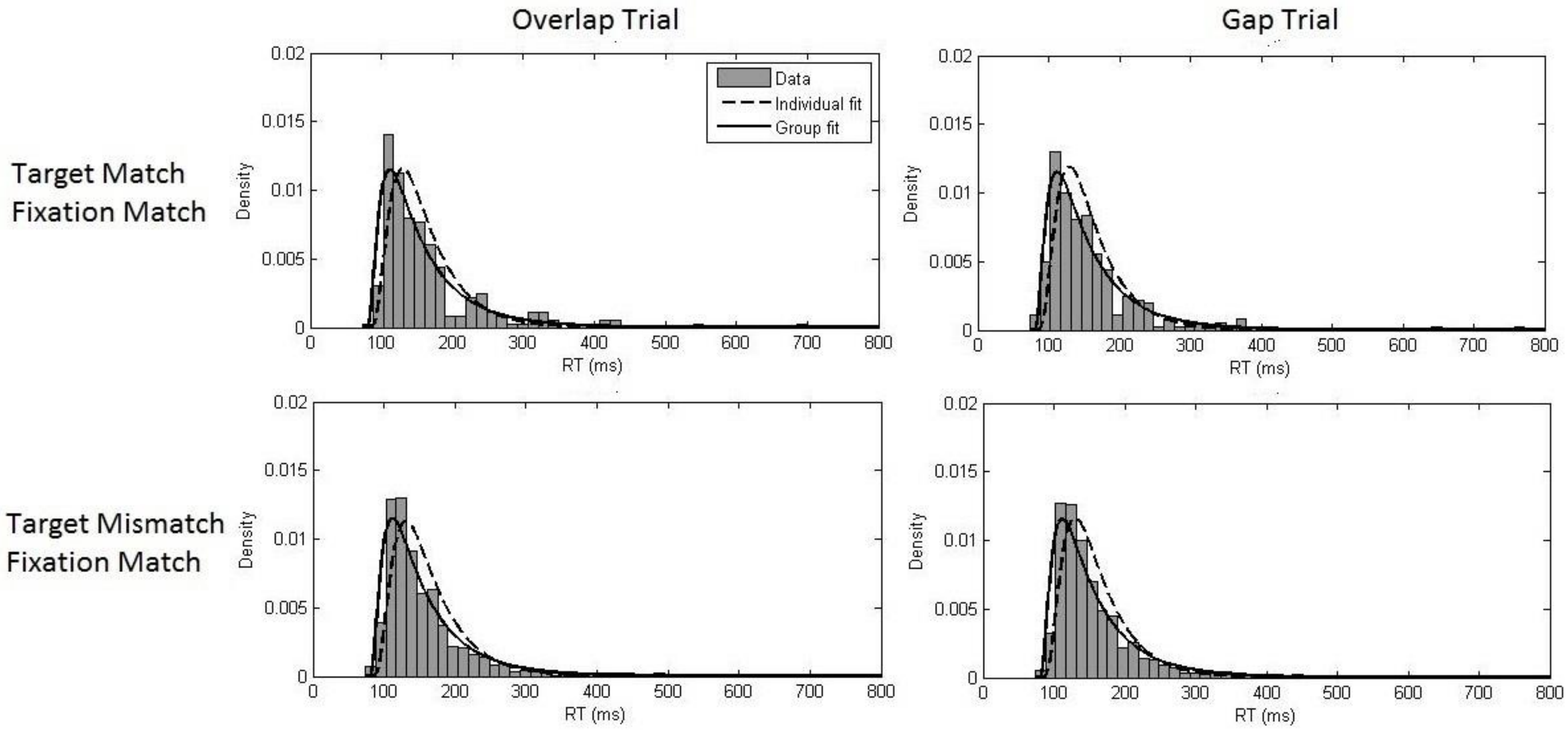
Figure 10. The predicted response time distributions for individual and group fits, and the observed data in Experiment 4. The predicted RT distributions for individual fits are dotted lines, and the predicted RT distributions for group fits are solid lines. The normalized histograms are the experimental data including all participants' trials.

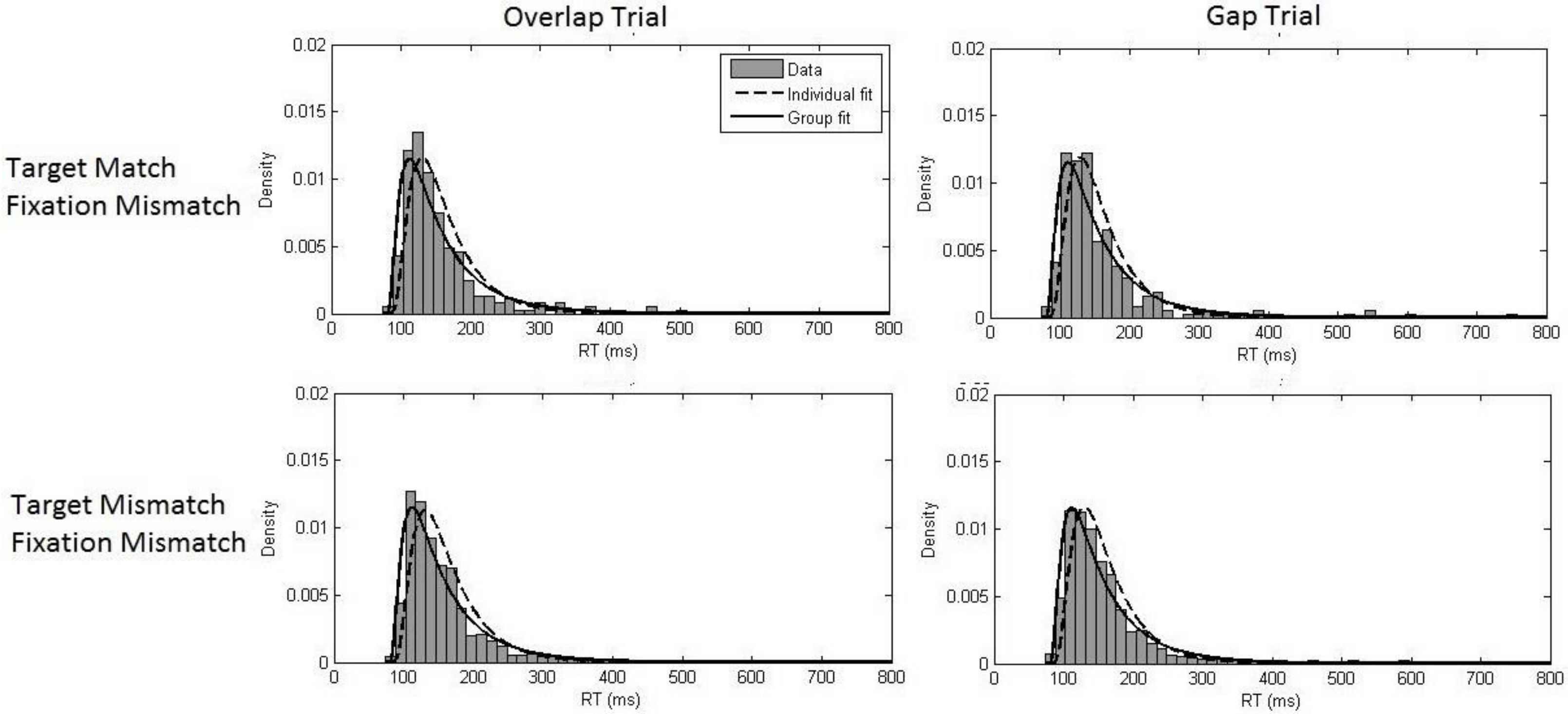
a. Mostly target match condition: Target match first group



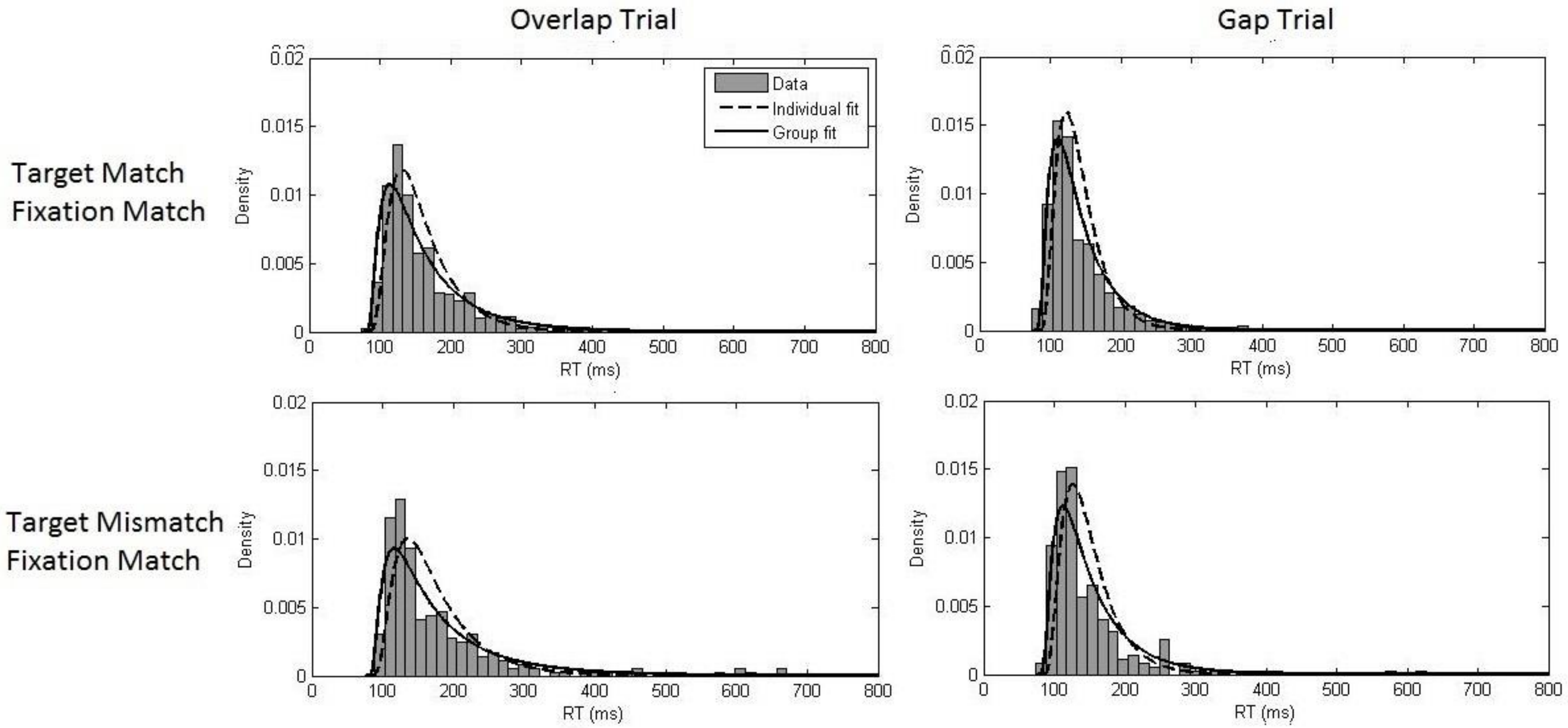

Overlap Trial
Gap Trial
Target Match
Fixation Mismatch
Target Mismatch
Fixation Mismatch
Data
Individual fit
Group fit
Density
0.02
0.015
0.01
0.005
0
0
100
200
300
400
500
600
700
800
RT (ms)

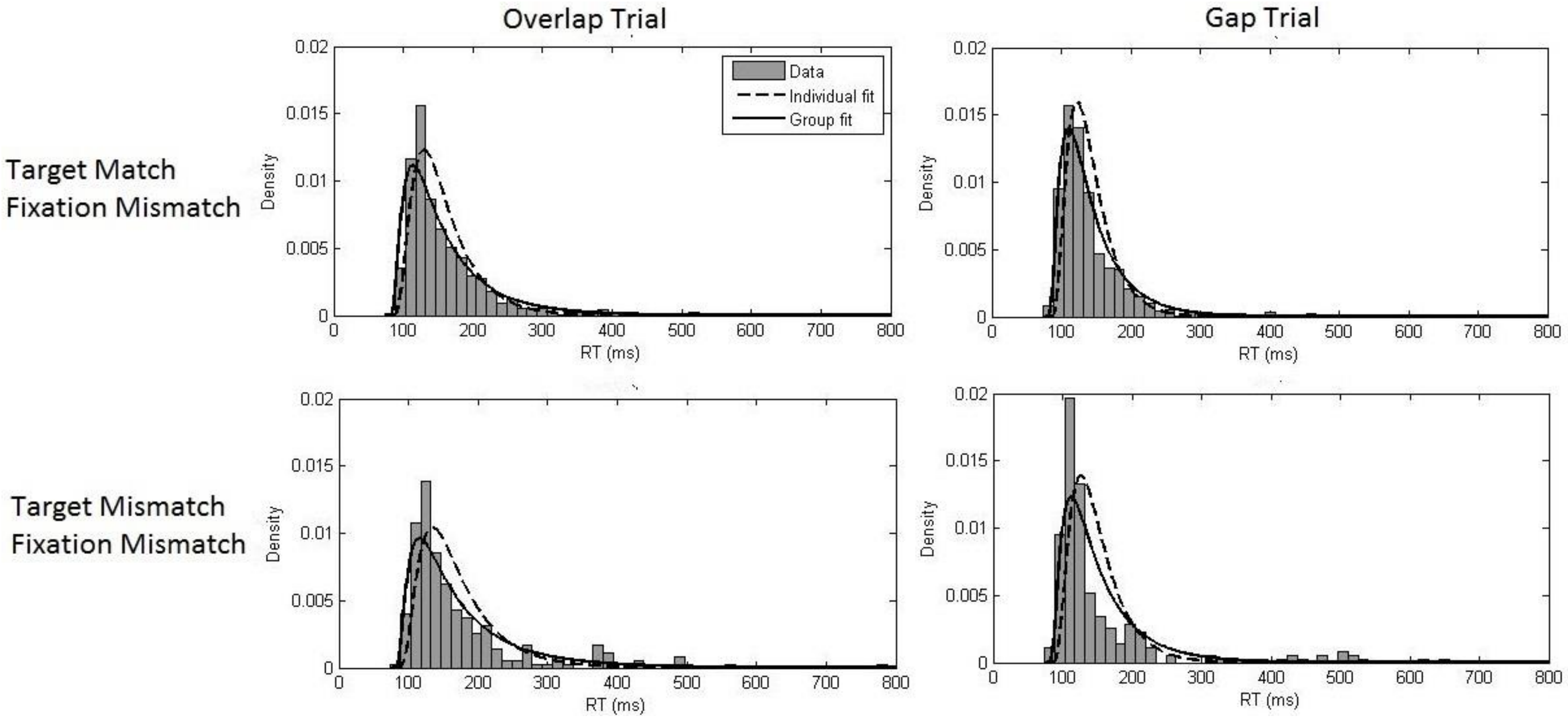
b. Mostly target mismatch condition; Target match first group



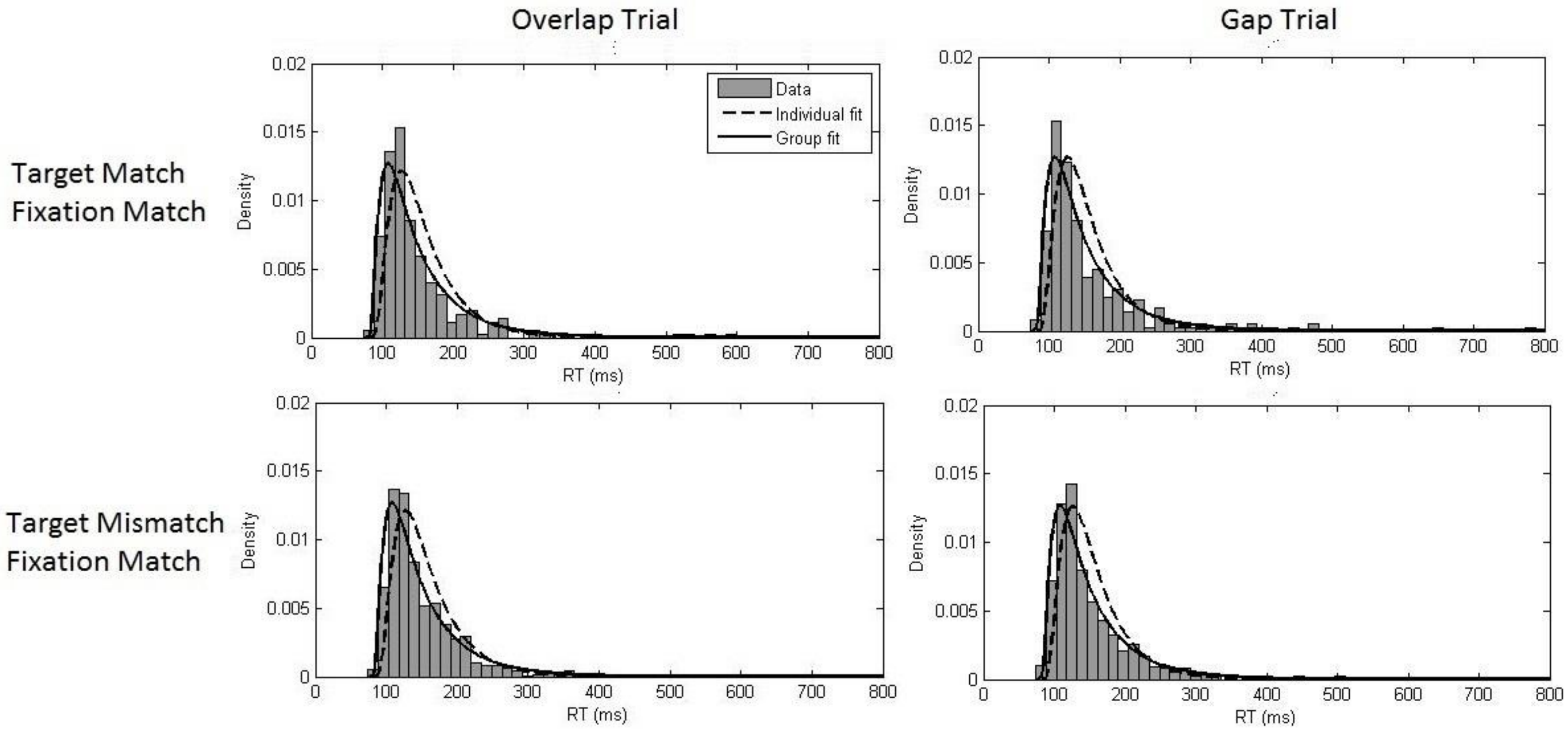

Overlap Trial
Gap Trial
Target Match
Fixation Mismatch
Target Mismatch
Fixation Mismatch
Data
Individual fit
Group fit
Density
0.02
0.015
0.01
0.005
0
RT (ms)
0
100
200
300
400
500
600
700
800

c. Mostly target match condition: Target mismatch first group




Overlap Trial
Gap Trial
Target Match
Fixation Mismatch
Target Mismatch
Fixation Mismatch
Data
Individual fit
Group fit
Density
RT (ms)
0
0.005
0.01
0.015
0.02
0
100
200
300
400
500
600
700
800

d. Mostly target mismatch condition: Target mismatch first group



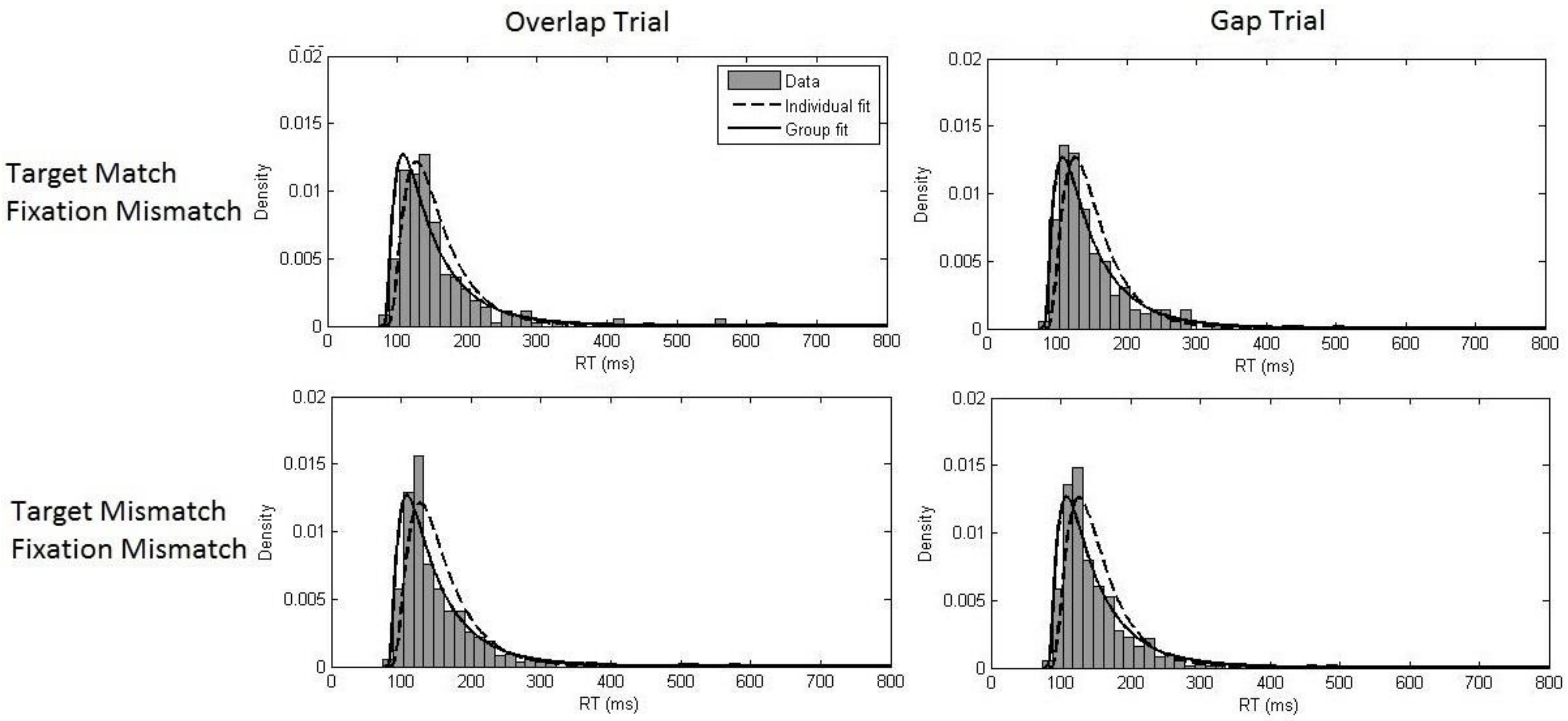
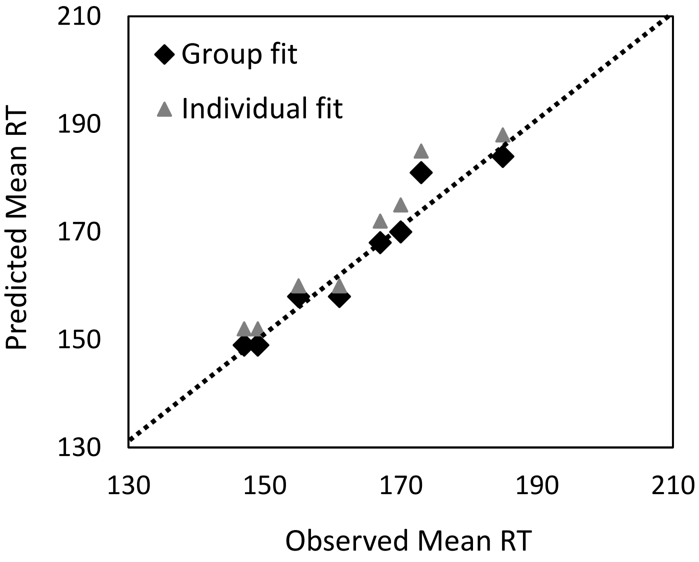
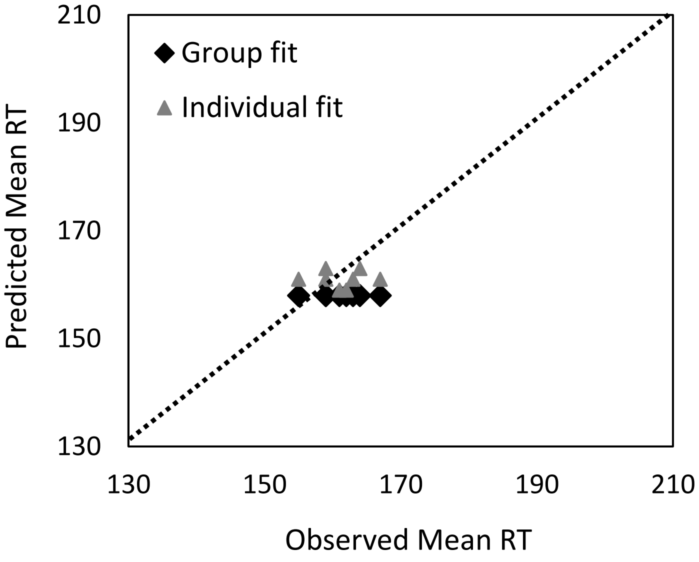

Overlap Trial
Gap Trial
Target Match
Fixation Mismatch
Target Mismatch
Fixation Mismatch
Data
Individual fit
Group fit
Density
0
0.005
0.01
0.015
0.02
RT (ms)
0
100
200
300
400
500
600
700
800

Figure 11. Mean response times (RTs) for the individual and group fits in Experiment 4. The dotted diagonal lines represent the perfect fit.

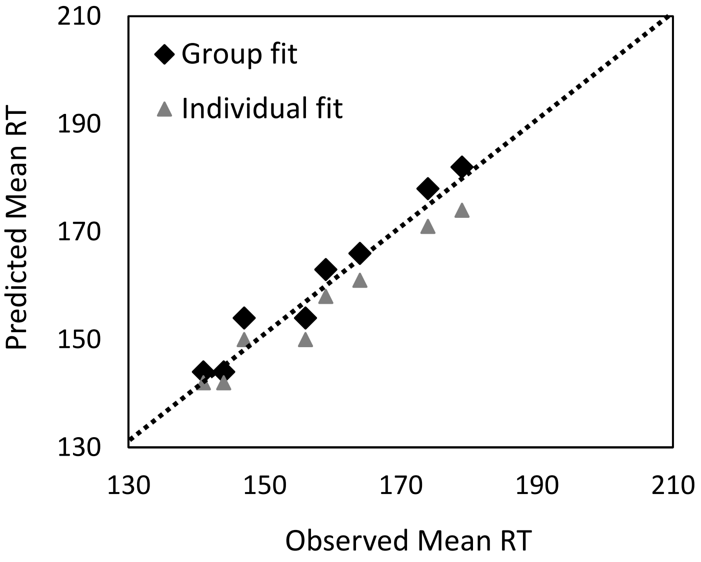
a. Mostly target match condition: Target match first group



b. Mostly target mismatch condition; Target match first group



c. Mostly target match condition: Target mismatch first group



d. Mostly target mismatch condition: Target mismatch first group

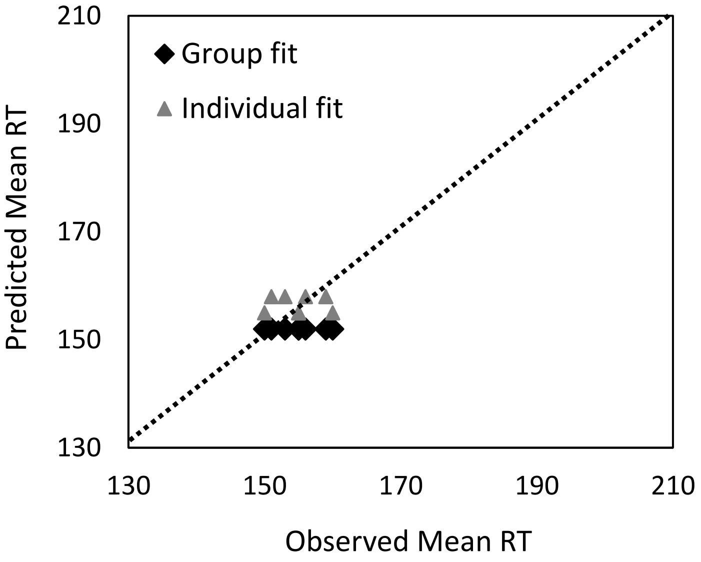
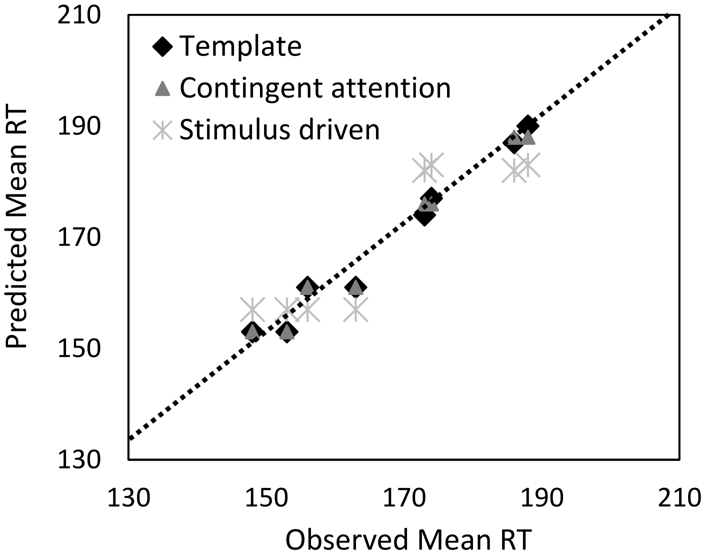


Figure 12. Mean response times (RTs) for group fits of the template model, the contingent attention account, and the purely stimulus driven account. The dotted diagonal line represents the perfect fit.



Appendix A

The LCA and Wald Diffusion Implementations of the Interactive Race Model

The interactive race model consists of two or more competing accumulators that are mutually inhibitory to each other (Boucher et al., 2007; Logan et al., 2015). The original interactive race model was implemented within the framework proposed by Usher and McClelland (2001), *leaky competing accumulator* (LCA). The LCA involves a stochastic differential equation (see Equation 1), and its implementation requires extensive simulation. This is computationally expensive, and a full implementation of the original interactive race model would require a high performance computing system, which is not available to many researchers. Therefore, in the present study, we sought an alternative implementation of the interactive race model that is mathematically more tractable and thus requires less computation to optimize model parameters. We have chosen to use the Wald diffusion model (see Logan et al., 2014), which has been used to model the stop-signal task that is similar to the countermanding task for which the original interactive race model was developed. Also, both models are part of the same class of models that assume diffusion processes to represent response activation. Nonetheless, the Wald diffusion model has a closed mathematical form (see Equation 2), and extensive simulation is not needed for its implementation. In this appendix, our purpose is to examine how well the Wald diffusion model mimics the original interactive race model. To this end, we first simulated the interactive race model with the LCA formulation for Experiment 3, and then we fit our template model to the simulated data to examine whether the simplification of the interactive race model had any consequence. To simulate the interactive race model, we used a model variation developed by Logan et al. (2015), which provided detailed dynamics of the competing accumulators that are responsive to the target (*go unit*) and to the fixation (*stop unit*), respectively.

Simulation of the Interactive Race Model

The interactive race model is depicted in Figure 4b. In a prosaccade task, participants maintain their gaze at a fixation mark at the start of a trial until the target appears on the screen. To maintain at the fixation, the stop unit has to continue inhibiting the go unit, so that the go unit activation is suppressed until the target appears. Upon onset of the target, the go unit gets activated and overcomes the inhibition from the stop unit. The go unit would be activated rapidly on a gap trial because an early offset of the fixation leads to decay of the stop unit activation before target onset, which weakens the inhibition of the go unit before the target appears. On an overlap trial, the fixation remains on the screen, and the stop unit continues inhibiting the go unit unless the go unit depresses the activation of the stop unit completely. Therefore, the drift rate (i.e., the speed of activation) of the go unit is generally faster on a gap trial than on an overlap trial, producing the gap effect (see Figure 5). Formally, the interactive race model is defined by the following two stochastic differential equations: for the go unit,

$$da_{go} = \frac{dt}{\tau}\left[v_{go} - \kappa a_{go}(t) - \beta_{stop} a_{stop}(t)\right] + \sqrt{\frac{dt}{\tau}}\xi, \qquad \text{(A1)}$$

and for the stop unit,

$$da_{stop} = \frac{dt}{\tau}\left[v_{stop} - \kappa a_{stop}(t) - \beta_{go} a_{go}(t)\right] + \sqrt{\frac{dt}{\tau}}\xi, \qquad \text{(A2)}$$

where a is the activation of the unit, v is the drift rate, κ is a constant representing the speed of decay of activation, and β is a constant representing lateral inhibition. The parameter τ determines the precision of the iteration, which is the unity (1 ms) in the present study, and ξ represents an independently and identically distributed random error across time, which is a Gaussian random variable with zero mean and the unit standard deviation. In the current formulation, v is constrained by the relation, $v = max(a) \cdot \kappa$, which guarantees that the activation of a unit has an asymptote. This assumption was necessary because the model assumes that the stop unit is asymptotically activated and continues inhibiting the go unit until the target occurs; otherwise, the stop unit activation increases infinitely, and the go unit would never be activated (see Logan et al., 2015). In addition to these parameters, the go unit also involves the response threshold θ; if the activation of the go unit exceeds the threshold, the process is terminated and a saccade is produced. The decision time is then defined by the minimum time, T, by which the go unit activation exceeds the threshold, plus the nondecision time t_0,

$$RT = \{min(T);\, a_{go} > \theta\} + t_0.$$

Note t_0 consisted of two components in the interactive race model, a delay of the accumulation process (D) and an arbitrary ballistic process after the go unit activation exceeds the threshold. The delay constant was set at the lower RT cutoff (80 ms) and the ballistic process was set at 10 ms (see Boucher et al., 2007).

To implement the interactive race model for Experiment 3 of the present study, we made assumptions similar to those of the template model, which we presented in the main text. In particular, when the short-term template matched the target color, there was an acceleration of attention shifting, which corresponded to the increased activation rate of the go unit. Similarly, when the short-term memory template matched the fixation color, there was an impairment of attention disengagement, which corresponded to the increased inhibition of the go unit. These two mechanisms were controlled by a parameter r, which multiplied the drift rate of the go unit on target match trials and the drift rate of the stop unit on fixation match trials. Therefore, the interactive race model for Experiment 3 required three parameters (v, κ, and β) for each of the go and stop units plus the response threshold θ and the accelerating parameter r, which was identical for the two units; other parameters were held constant. As noted earlier, a full implementation of the model requires high computational power, and we had no access to such a powerful system. Therefore, the model was not fit to the actual experimental data; instead, we set the parameters arbitrarily to produce a pattern of the data that were similar to the results of Experiment 3. These parameters are shown in Table A1, and the simulated RTs are summarized in Figure A1a. The simulation was run 1000 trials for each of the eight conditions, and the template model (Equation 2) was fit to the simulated data. The fitting procedure was identical with that used for the actual experimental data of Experiment 3.

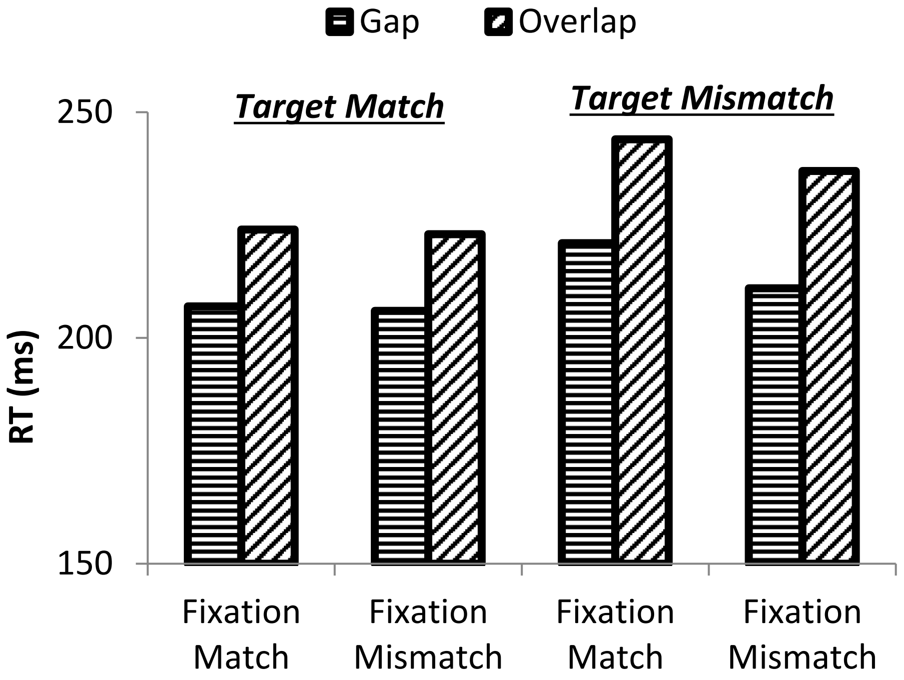
The best parameters of the template model is also shown in Table A1. It is not meaningful to compare these parameters to those obtained for the actual experimental data (see Table 2) because they were fitted to different data. Instead, we simply note that the results showed an excellent fit of the model to the interactive race model (R^2 = .941). The template model somehow underestimated mean RTs (see Figure A1b), but there were little deviations of the predicted RT distributions from the simulation of the interactive race model (see Figure 2A). Although the Wald diffusion model is markedly easier to fit to the data than is the original interactive race model, we do not wish to make a strong claim as to whether our approach should replace the original interactive race model. Such a claim would require more extensive comparisons of the two approaches, ideally by fitting the interactive race model to the same experimental data as the Wald diffusion model. Therefore, we only suggest that the Wald diffusion model mimics the LCA implementation, making it a viable alternative implementation of the interactive race model.

Table A1. The simulation parameters of the interactive race model, and the best parameters of the template model fit and the goodness of fit measures (R^2 *and sum of squared errors [SSE]*).

Interactive race model			Template model	
	Go unit	Stop unit		
v	.1500	.0375	a	.112
κ	.0100	.0025	i	.013
β	.0100	.0080	z	18.356
D	80	80	t_0	53.807
r	1.100	1.100	r	1.076
θ	12	-	-	-
			R^2	.941
			SSE	510

Figure A1. Mean response times (RTs) for the interactive race model simulation (a) and the template model fit (b).

(a)



(b)

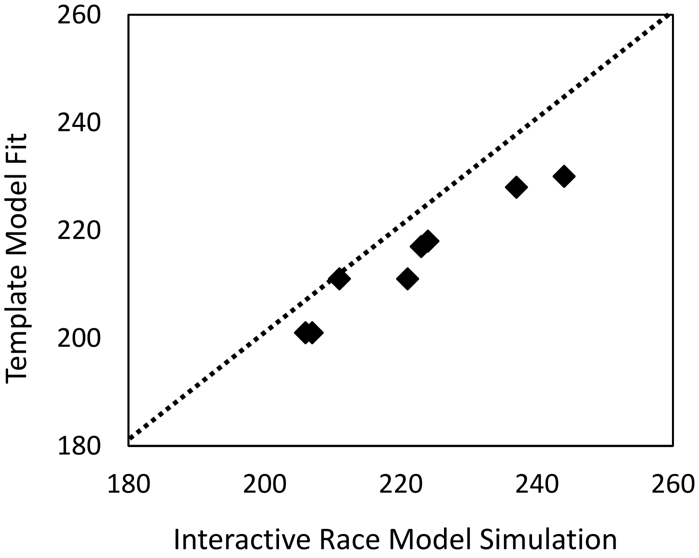


Figure A2. The simulated response time distributions of the interactive race model (histograms) and the predicted response time distribution of the template model (dotted lines).