A Dual-Stage Two-Phase Model of Selective Attention

Ronald Hübner, Marco Steinhauser, and Carola Lehle
Universität Konstanz

The dual-stage two-phase (DSTP) model is introduced as a formal and general model of selective attention that includes both an early and a late stage of stimulus selection. Whereas at the early stage information is selected by perceptual filters whose selectivity is relatively limited, at the late stage stimuli are selected more efficiently on a categorical basis. Consequently, selectivity is first low but then abruptly increases during the course of stimulus processing. Although intended as a general model of selective attention, in the present study the DSTP model was applied to account for the distributional data of 3 flanker task experiments. The fit of the model to the data was not only rather good but also superior to those of alternative single-stage models with a continuously increasing selectivity. All together, the model provides a comprehensive account of how early and late stages of attention interact in the control of performance.

Keywords: selective attention, early selection, late selection, flanker task, diffusion process

A characteristic of human behavior is that it is controlled by internal as well as external factors and that these factors interact in complex ways. For instance, it is well known that stimuli in the environment can automatically control behavior by activating associated responses (cf. Desimone & Duncan, 1995; Kinchla, 1992; Pashler, Johnston, & Ruthruff, 2001). Although such an external control can be highly efficient, it might be inappropriate in situations where externally activated responses are not in accord with the current goal. Therefore, in addition to external control mechanisms, internal control mechanisms are also needed to ensure that goal-relevant stimuli determine the behavior, particularly in situations with ambiguous stimuli (cf. Logan, 1980; Norman & Shallice, 1986; W. Schneider, Dumais, & Shiffrin, 1984).

One important internal control mechanism for this objective is selective attention. Accordingly, numerous theories have been proposed to account for the ability to process stimuli selectively. Originally, there was a dispute about whether selection takes place early or late in the course of stimulus processing (cf. Kahneman & Treisman, 1984). *Early selection* accounts assumed that stimulus selection occurs early and is based on elementary stimulus features that are available at preidentification stages (cf. Broadbent, 1958; Johnston & Dark, 1982; Neisser, 1976). According to this view, only one object at a time is selected for identification and further processing. In contrast, *late selection* accounts assumed that several stimuli can be identified in parallel, and a target stimulus is

then selected for further processing at a later stage based on identity or other semantic features (cf. Deutsch & Deutsch, 1963; Duncan, 1980; Moray, 1959; Shiffrin & Schneider, 1977).

However, there are also theories assuming that early and late selection represent two modes of attention, which are applied depending on their costs and benefits (e.g., Johnston & Heinz, 1978). Recent evidence supports these theories by showing that modes of selective attention can be chosen strategically (Lehle & Hübner, 2008) or are induced by current task demands (de Fockert, Rees, Frith, & Lavie, 2001; Lavie, Hirst, De Fockert, & Viding, 2004). Despite these ideas, though, many questions remain open. For instance, it is still unclear whether early and late selection mechanisms operate simultaneously or serially and how exactly these mechanisms interact to enable flexible and coherent behavior.

One source of information that has been used to investigate these questions is patterns of interference effects in distributional data from conflict paradigms such as the flanker task (B. A. Eriksen & Eriksen, 1974) or the Stroop task (Stroop, 1935). These patterns clearly indicate an increase in selectivity during the course of processing. Whereas some researchers interpreted these patterns as evidence for a less selective early stage of processing followed by a more selective late stage (e.g., Gratton, Coles, & Donchin, 1992), current formal models account for these results by assuming a continuously increasing selectivity of a single selection process (e.g., Cohen, Servan-Schreiber, & McClelland, 1992; Heitz & Engle, 2007; Liu, Yu, & Holmes, 2009; Spieler, Balota, & Faust, 2000; Yu, Dayan, & Cohen, 2009).

In the present article we present a formal model that sticks to the idea of an early and a late selection stage by assuming a discrete transition from a less selective to a more selective stage. The core assumption of our *dual-stage two-phase* (DSTP) model is that competition between early and late selection processes results in two consecutive phases of response selection and that these phases differ with respect to their susceptibility to interference. Although early and late selection operate simultaneously, they contribute

differentially to performance, depending on the relative efficiency of each process.

Although our model is sufficiently abstract to potentially serve as a framework for interpreting distributional effects in a large range of conflict paradigms, in this article we use the DSTP model to account for the performance in the flanker task, a spatial selection task that has become a standard paradigm for investigating selective attention. A great advantage of the flanker task is that the difficulty of stimulus selection can systematically and easily be manipulated. We show that the DSTP model can not only account for the observed pattern of interference in distributional data but also that it is superior to models assuming a continuously increasing selectivity. However, before we report our experiments and model fits, we first consider results and concepts in the area of spatial visual attention that provide further motivation for the DSTP model and then introduce our model in more detail.

Spatial Visual Attention

An essential mechanism for early selection in the visual domain is spatial filtering. Several metaphors have been used to describe the way it operates. The *attentional spotlight* metaphor assumes that visual attention can be allocated to a certain location in the visual field and that items at that location are processed more intensively than items at other locations (Posner, 1980; Posner, Snyder, & Davidson, 1980). The *attentional zoom-lens* metaphor generalizes this idea of a spotlight by assuming that not only is the position of the attentional filter variable but also, at least within certain limits, that its size and form are (B. A. Eriksen & Eriksen, 1974; C. W. Eriksen & Hoffman, 1973; C. W. Eriksen & Schultz, 1979; C. W. Eriksen & St. James, 1986). Furthermore, it has been assumed that changes of the distribution of spatial attention can occur abruptly rather than continuously (Reeves & Sperling, 1986; Shih & Sperling, 2002; Sperling & Weichselgartner, 1995).

Important properties of spatial attention have been revealed by the flanker task (B. A. Eriksen & Eriksen, 1974). In this task participants have to identify a target stimulus as fast and as accurately as possible, while ignoring irrelevant flanker stimuli. The flankers are usually congruent, that is, associated with the same response as the target, or incongruent, that is, associated with the opposite response. The extent to which the flankers can be ignored is assessed by the difference between the performance for congruent and incongruent stimuli, which is called the *flanker congruency effect*. Usually, responses to congruent stimuli are faster and more reliable than responses to incongruent flankers, and the size of differences in response time (RT) and error rate are considered measures of the efficiency of selective attention.

The flanker task and its variants have been used to examine the possible forms of the spatial filter (e.g., LaBerge & Brown, 1986; M. M. Müller & Hübner, 2002; N. G. Müller, Mollenhauer, Rösler, & Kleinschmidt, 2005) and whether its area is contiguous (e.g., Franconeri, Alvarez, & Enns, 2007; Kramer & Hahn, 1995; McMains & Somers, 2004; M. M. Müller, Malinowski, Gruber, & Hillyard, 2003). LaBerge and Brown (1989), for instance, proposed *attentional gradients* to account for gradual changes of attentional resources across the visual field (see also LaBerge, Brown, Carter, & Bash, 1991; Logan, 1996). These results suggest that the distribution of spatial attention in the visual field is rather flexible. Indeed, Fazl, Grossberg, and Mingolla (2009) recently

introduced the concept of *attentional shrouds* (see also Tyler & Kontsevich, 1995) that can even fit to an object's form in order to guide object recognition.

Spatial selection is also part of some formal models of visual attention. The theory of visual attention (TVA; Bundesen, 1990), for instance, has been applied to numerous empirical phenomena (for an overview see Logan, 2004). It consists of a *perceptual filtering* mechanism that differentially weights spatial locations to allow spatial selection in multi-item displays. Additionally, TVA has a categorical biasing mechanism, which is called *pigeonholing*, and which is thought to bias the selected items with respect to the semantic categories that are relevant for the task. Filtering and pigeonholing loosely correspond to early and late selection mechanisms, respectively. However, the crucial difference to traditional views is that they are considered two aspects of the same process rather than two different stages of processing (see also the discussion of this issue in Logan, 2002). Because perceptual biasing and categorical biasing affect processing at the same time, TVA can be considered a single-stage model.

Although TVA can account for various phenomena such as spatial cuing effects (Posner et al., 1980), it cannot explain attentional selection in conflict paradigms. With respect to the flanker task, for instance, it correctly predicts a congruency effect in the error rates but fails to predict such an effect in RT. Therefore, Logan (1996) extended TVA and added a contour detection system (Compton & Logan, 1993) for representing space and objects more explicitly. With these extensions the model could not only account for the general congruency effects in the flanker task but also for effects of the distance between target and flankers (see also Logan, 2002; Logan & Gordon, 2001).

Phenomena like the congruency effect show that although perceptual filtering is an important early selection mechanism, it is not perfect. This raises the question of what happens if early selection fails or cannot be applied. Evidence suggests that late selection mechanisms come into play in this case. This is in accord with variable-stage models, which assume that relevant stimuli can be selected at a variable stage of processing, depending on the task demands (e.g., Pashler & Badgio, 1985; Yantis & Johnston, 1990). Although assuming more than one stage of selection seems plausible, variable-stage models are difficult to validate, especially if one takes only mean RTs and mean error rates into account.

One way to obtain more detailed information about the dynamics and flexibility of stimulus selection is to consider distributional data. For instance, if one compares the RT distributions for correct responses between congruent and incongruent flanker stimuli, then one does not only find that the distribution for incongruent stimuli is shifted toward slower responses but also that the congruency effect increases with RT (e.g., Ridderinkhof, 2002), that is, that it gets larger for slower responses (but see B. A. Eriksen, Eriksen, & Hoffman, 1986; Spieler et al., 2000). In our experiments reported below, we see further examples of this phenomenon. If one analyzes accuracy, though, then one usually finds that the congruency effect decreases with RT. A method that has been used to demonstrate the latter effect is to show so-called *conditional accuracy functions*. Empirically, these functions are constructed by calculating the accuracy for different RT bins or quantiles.

Gratton, Coles, Sirevaag, Eriksen, and Donchin (1988) examined conditional accuracy functions for the performance in the flanker task and observed for congruent stimuli (except for very

fast responses which were presumably due to fast guesses) that accuracy was already high for fast responses. In contrast, the accuracy for incongruent stimuli was rather low for fast responses but then improved quickly with RT up to the accuracy level of the congruent stimuli. Because conditional accuracy functions are completely determined by the RT distributions of correct and incorrect responses and the corresponding response probabilities (cf. Luce, 1986), their characteristics can also be interpreted in terms of distributional properties. Flat conditional accuracy functions, for instance, are the result of similar RT distributions for correct and incorrect responses. In contrast, increasing conditional accuracy functions indicate that error responses tend to be faster than correct responses. Indeed, the typical observation from the flanker task that incongruent stimuli but not congruent ones produce an increasing conditional accuracy function results from the fact that, whereas correct responses are slower for incongruent stimuli than for congruent stimuli, error responses are similarly fast for both stimulus types or even faster for incongruent stimuli.

The observation that the congruency effect differs in size between different parts of RT distributions led to the idea that fast responses are produced by different processes than slow responses. Gratton and his colleagues (Gratton et al., 1992; Gratton et al., 1988), for instance, proposed that an early and unselective phase of analysis is followed by a selective phase (see also Coles, Gratton, Bashore, Eriksen, & Donchin, 1985). Responses that occur during the first phase show a strong congruency effect because all stimuli in the display activate their associated response during this phase. In contrast, nearly no congruency effect is obtained for responses that occur during the second phase because spatial attention is focused on the target during this phase. Obviously, the proposed two phases are similar to the parallel phase and serial phase in visual-search models (e.g., Hoffman, 1978; Logan, Withey, & Cowan, 1977; Treisman, 1988; Treisman & Gelade, 1980; Wolfe, 1994).

Whereas Gratton and colleagues suggested that the early phase is entirely unselective, subsequent attempts to model distributional flanker task data (e.g., Cohen et al., 1992) revealed that it is necessary to assume that already initial stimulus processing must be selective to some extent. As we see later in this article, our investigation supports this conjecture. Moreover, all subsequent formal models abandoned the idea of discrete phases and assumed that selectivity increases continuously within a single stage (Cohen et al., 1992; Liu, Holmes, & Cohen, 2008; Liu et al., 2009; Yu et al., 2009). In contrast, our DSTP model is the first formal model that accounts for the distributional data by assuming discrete phases. As we show, discrete phases can better account for the data than a single stage with an increasing selectivity.

Another class of models that can potentially account for the flanker task data are dual-route models (e.g., De Jong, Liang, & Lauber, 1994; Hommel, 1998; Kornblum, Hasbroucq, & Osman, 1990; Ridderinkhof, 1997), which were originally developed to describe automatic versus controlled processing (cf. Logan, 1988; W. Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). These models assume that stimulus information can affect response selection via a fast automatic route and via a slow controlled route. Through the automatic route both task-relevant and task-irrelevant stimulus information directly activate the involved response units according to learned stimulus–response associations. By the controlled route, in contrast, stimulus information is

deliberately transformed into response activation according to the stimulus–response mapping defined for the current task. Dual-route models can explain the low accuracy for fast responses to incongruent flanker stimuli by assuming that flankers activate their associated response units via the fast automatic route. If one further assumes that the slower responses are produced mainly by the controlled route, which allows only the target to activate its response unit, then this accounts for the low error rate for slow responses.

Obviously, the considered results in the area of spatial visual attention suggest that even performing simple tasks involves at least two stages, routes, or phases of stimulus processing, whose output is differentially reflected by fast and slow responses. Current formal models, though, which assume only a single stage, or additionally a continuously increasing selectivity, cannot appropriately account for these results, as is shown in some detail below. Therefore, to provide a better alternative, we developed our DSTP model, which is introduced in the next section. It extends and integrates valuable ideas from previous variable-stage, dual-phase, and dual-route models.

The DSTP Model

The main characteristics of the DSTP model are two discrete stages of stimulus selection, an early stage and a late stage. The early stage selects stimuli by perceptual filtering and is not strictly selective in the sense that information from irrelevant stimuli (e.g., flankers) might also be selected to a certain degree. This stage provides information for response selection during a first phase. At the late stage information is selected from only one stimulus (e.g., the target) based on its identity, which is highly selective. The selected stimulus then drives response selection during a second phase of response selection (see Figure 1). We further assume that selective attention serves in the first place for the selection of task-related responses. Therefore, a central part of our model is a response selection mechanism.

As we are concerned with both the selection of stimuli and the selection of responses, it is important not to confuse the different terms. Thus, for clarity, we use the terms *early* and *late* exclusively in connection with stimulus selection. For the phases of response selection we use the attributes *first* and *second* or the terms *Phase 1* and *Phase 2*. We now describe these mechanisms and concepts in more detail and how we formalized our model.

In the DSTP model, response selection proceeds by a diffusion process (cf. Ratcliff, 1978; Ratcliff & Rouder, 1998; Smith, 2000), which accumulates evidence in favor of the one or the other response. The evidence results from a sequence of noisy samples extracted from the presented stimuli. These samples are accumulated over time and the corresponding evidence is represented by the state $X(t)$, $t \geq 0$ of a stochastic process. It is assumed that the samples are normally distributed with mean μ, which is called the *rate* or *drift* of the process, and variance σ^2, which is called the *diffusion coefficient* of the process. In our case, the drift depends on the stimulus conditions and on the characteristics of the observer (e.g., on its attentional capacity).

In order to obtain a response, decision boundaries A and $-B$ are defined, which represent the two possible Responses A and B, respectively. If the diffusion process, after starting from state $X(0)$, hits one of the boundaries, the corresponding response is selected
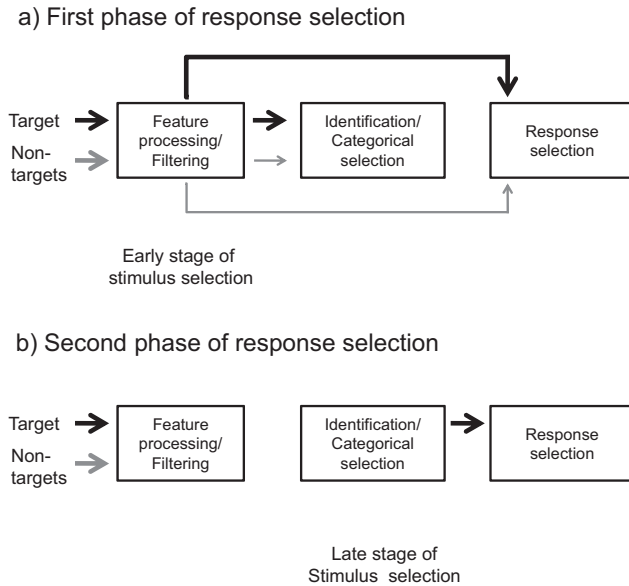
a) First phase of response selection

Early stage of
stimulus selection

b) Second phase of response selection

Late stage of
Stimulus selection

Figure 1. Flow diagrams for the two phases of the dual-stage two-phase (DSTP) model. Panel a: During the first phase, stimulus information passing the perceptual filter not only serves for stimulus identification but also affects the response selection process. Because perceptual filtering is imperfect, nontargets (gray arrows) also have some effect. Panel b: In the second phase, that is, after stimulus identification and selection, response selection is driven exclusively by the selected stimulus (ideally the target).

(and also executed). Because we are concerned only with symmetric boundaries, we also use the term *criterion*, which simply represents the distance from zero to each boundary and in this case can be represented by A alone. The time from the start of the process until it hits a boundary is a random variable T and called *first passage time*. It usually reflects the time used for response selection or the decision time, which is an important component of the overall RT.

In the DSTP model, we distinguish two phases of response selection, whose processes are denoted by $RS1$ and $RS2$. In the first phase $RS1$ is driven by the output of an early stage of stimulus selection, where basic stimulus information is selected by perceptual filters (see Figure 1a). As we are concerned with multiple-item displays, we assume that the information provided by each item in the display is summed up (cf. Diederich & Colonius, 1991; Logan, 1980, 1996; Logan & Gordon, 2001; D. W. Schneider & Logan, 2005; W. Schwarz, 1989). Perceptual filtering (see the boxes on the far left in Figures 1a and 1b) is taken into account by assuming that the individual components of the overall rate are already the product of the evidence provided by the corresponding item and an attentional weight (cf. Bundesen, 1990; Logan, 1996). For the flanker task, for instance, we assume component rates μ_{ta} and μ_{fl} for target and flankers, respectively. Because each item provides the same perceptual evidence for its category membership (or response), irrespective of whether it functions as target or as flanker, it can be represented mathematically by a constant with the same value for each possible item on each position. With this assumption, the component rates are products of attentional weights and the same constant. Consequently, observed variations in μ can be interpreted directly as variation in attentional weight.

If a response is already selected by $RS1$, it is executed and the trial is finished. However, because the efficiency of the perceptual filter is limited, some information from irrelevant items (e.g., flankers) usually also contributes to the overall rate (as indicated in Figure 1b by the long thin gray line), which increases the probability of an error. Thus, to obtain a higher accuracy, a late stimulus selection process SS, which is also implemented as a diffusion process, runs in parallel with $RS1$ (cf. Usher, Olami, & McClelland, 2002). In the flowchart of our model (see Figure 1) the late stimulus selection process for both the first and second phases is represented by the box in the middle. The late stimulus selection process selects the Target C or the Nontarget D when it hits the boundaries C and $-D$, respectively. If this occurs before a response is selected, then, from that point onward, response selection enters Phase 2 (see Figure 1b), where response selection proceeds by $RS2$, which is driven exclusively by the selected item. Thus, although SS starts with stimulus onset, because its output is discrete and occurs at a relatively late point in time, it is justified to characterize the corresponding selection stage as being late.

In Figure 1 the input into the late stimulus selection stage also passed through the perceptual filter. Although this is plausible, in the formal version of our model we do not assume such a dependency. Rather, the input into SS is represented by the specific value of the rate parameter μ_{SS}.

Together, our model's processing of the flanker task can be categorized into six possible situations, which are shown in Figure 2. The processes are arranged vertically, which illustrates more clearly the way that stimulus selection affects response selection. There are the two cases where a response is already selected before a stimulus is selected. In one case, $RS1$ selects Response A (see Panel a), and in the other case, it selects Response B (see Panel b). In the other four cases stimulus selection is finished before a response is selected, that is, SS wins the race and either selects the Target C (see Panels c and d) or a Nontarget D (see Panels e and f). When SS wins the race against $RS1$ by selecting a stimulus at time t_s, then $RS1$ will be in a certain state $X(t_s)$ somewhere between its boundaries of A and $-B$. $RS2$ continues from this state. Because the selected item determines the rate of $RS2$, the rate can be different in Phase 2, compared to Phase 1. In Figure 2 this is indicated by changes in the gray level for the corresponding pathways of $RS2$. Here, we simply assume that the rate of $RS2$ is μ_{RS2}, when the target or a congruent nontarget has been selected, and $-\mu_{RS2}$, when an incongruent nontarget has been selected. Finally, $RS2$ selects either Response A or B. As can also be seen in Figure 2, each response is possible irrespective of the chosen item. However, they occur with different probabilities. By combining the characteristics of the different events, the overall performance of the model can be computed (for the mathematical details of our model see Appendix A).

One strength of the DSTP model is its late selection mechanism that should account for the relatively reliable performance even in situations where early selection is poor. To see whether this is indeed the case, we conducted three flanker task experiments (reported in the next three sections) in which we manipulated the efficiency of early selection. In the first experiment we did this by varying the spacing between target and flankers, in the second experiment we manipulated the spatial uncertainty and eccentricity of the stimuli, and in the third experiment we varied the proportion of congruent and incongruent stimuli. For each experiment, we
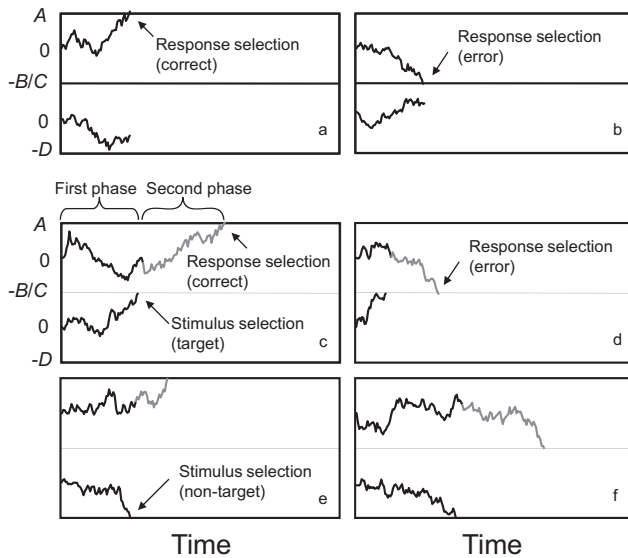
Figure 2. The possible situations that can occur during processing in our dual-stage two-phase (DSTP) model are shown. The upper curve in each panel shows an example path of the response selection process. Response A or B is selected when the state of the process hits boundary A or $-B$, respectively. The lower curve in each panel shows a sample path of the stimulus selection process. Target C or Nontarget D is selected when the state of the process hits boundary C or $-D$, respectively. Panels a and b represent examples where a response is already selected before stimulus selection ends. The other four panels show situations when there are two phases of response selection. Panel c, for instance, represents the situation where a target is selected, which changes the rate of the corresponding response selection process. The two phases of response selection are indicated by two different gray levels.

analyzed parameters of the DSTP model after fitting the model to the distributional data of each condition. We expected that late selection should be involved more strongly under difficult selection conditions and that this should be reflected by the corresponding parameter values of our model. After describing data and model fits for the three experiments, we compare the performance of the DSTP model with several alternative models.

Experiment 1: Spacing Between Target and Flankers

In our first experiment we tried to manipulate the relative contributions of early and late stimulus selection to performance by varying the spacing between target and flankers. It has been shown that the narrower the spacing, the larger the flanker congruency effect (e.g., B. A. Eriksen & Eriksen, 1974; Miller, 1991). Besides the hypothesis that the minimal diameter of the attentional zoom lens is limited (C. W. Eriksen & St. James, 1986; LaBerge et al., 1991), one possible reason for this phenomenon is that the spatial selection of single items from the display competes with the tendency of the visual system to construct larger perceptual units according to Gestalt principles (Wertheimer, 1922). Thus, if the flankers are close to the target, all items are strongly grouped by proximity (Kramer & Jacobson, 1991; Logan, 1996), which makes target selection difficult. Over the years it has been shown that other Gestalt principles, such as grouping by similarity (Baylis &

Driver, 1992; Harms & Bundesen, 1983; Hübner & Backer, 1999; Kramer & Jacobson, 1991) or grouping by common fate (Driver & Baylis, 1989), produce a similar limitation on target selection.

These results demonstrate that spatial filtering is not very efficient if flankers are located close to the target. Therefore, it can be assumed that, compared to a wider spacing, late selection is involved more strongly in this case. For instance, target selection could rely more on categorical information than on metric spatial information (see also P. Brown & Fera, 1994; LaBerge & Brown, 1989). To see how early and late selection vary with spacing, we conducted a flanker task experiment where the participants had to categorize numerals as odd or even. Each target numeral was flanked by two copies of a numeral that had either the same parity as the target (congruent condition) or the opposite parity (incongruent condition). In a narrow spacing condition the flankers were presented close to the target, whereas in a wide spacing condition the flankers occurred farther apart (details of the method are given in Appendix B). The mean results (see Appendix B) show that, as expected, the flanker congruency effect in RT and in the error rates was significantly larger in the narrow condition than in the wide condition.

To analyze and model the details of the data, we also considered the RT distributions. Following Ratcliff and his colleagues (e.g., Ratcliff & Smith, 2004), we represented the distributions for each participant by five quantiles (.1, .3, .5, .7, and .9). This was done for correct responses and error responses to congruent and incongruent stimuli, respectively. These quantiles were then averaged across participants to obtain group distributions.

Because response probabilities (i.e., accuracy) are also highly informative for the present objective, we depict our data by using so-called *defective* cumulative distributions (S. D. Brown & Heathcote, 2008; Feller, 1968), which refer to cumulative distributions weighted by the relative proportion of correct and incorrect responses. Accordingly, the distributions sum up to the rate of correct and incorrect responses rather than to 1. Figure 3 shows the graphs for the present experiment. If we consider the distributions for congruent stimuli in the wide condition, for instance, then we see that the 90% quantile for correct responses is graphed at a probability of .83, whereas that for incorrect responses is graphed at a probability of .07. This means that 90% of the 92% (i.e., 83%) correct responses were faster than 576 ms, whereas 90% of the 8% (i.e., 7%) incorrect responses were faster than 526 ms.

As can be seen in Figure 3, the narrow condition showed a generally increased flanker congruency effect in RTs as well as in accuracy. However, more important for our further considerations is the question of how these effects develop as a function of RT. On the one hand, the flanker congruency effect in RTs for correct responses is increasing with RT. This can be seen by considering the horizontal distance between corresponding data points for the congruent and incongruent conditions. This distance increases with the height of the quantile. On the other hand, the flanker congruency effect in accuracy is largest for lower and intermediate quantiles. This can be seen by considering the slopes of the distributions for incorrect responses. For lower and intermediate quantiles, the slope is greater for the incongruent than for the congruent condition, indicating that more fast errors occurred for incongruent stimuli. For this reason, the distribution of incorrect responses to incongruent stimuli is compressed so that higher quantiles have faster RTs for incongruent stimuli than for congru-
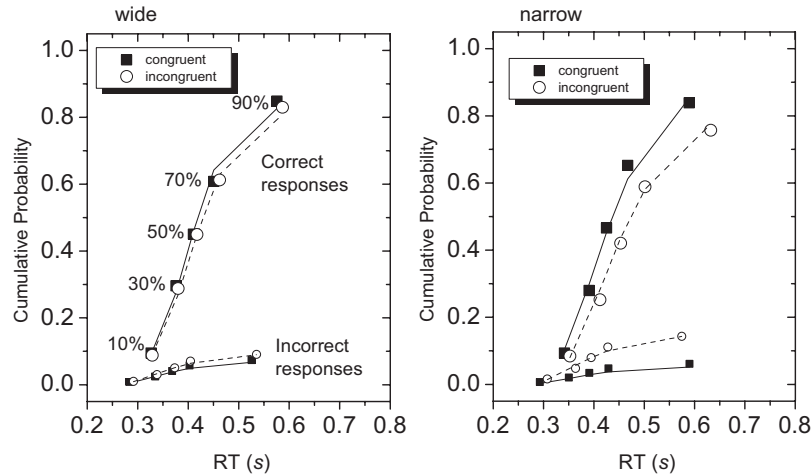
Figure 3. The two panels show defective cumulative response time (RT) distributions (quantile averaged) separately for congruent and incongruent trials of the two spacing conditions (wide vs. narrow) in our first experiment. The two upper distributions in each panel refer to correct responses, whereas the two lower distributions in each panel refer to error responses. Symbols represent experimental data and lines represent fit values of the dual-stage two-phase (DSTP) model.

ent stimuli. This is also reflected by the observation that the mean error RT was shorter for incongruent than for congruent stimuli (416 ms vs. 438 ms).

All together, these effects amount to the known phenomenon that the flanker congruency effect in RT is larger for slow than for fast responses, whereas the effect in the error rates is mainly present for fast responses. As mentioned, this difference, which is particularly large under high flanker interference, can be interpreted as an indication that two stimulus selection processes are involved. The fast but unreliable early selection process is responsible for the congruency effect in the error rates, whereas the slow but reliable late selection process produces the effect in RT. To see whether this interpretation is also reflected by the corresponding parameter values of the DSTP model, we fit the model to the distributional data.

Model Fit

As mentioned, we assumed that the overall rate μ of the response selection process (*RS*1) in the first phase is the sum of the contribution μ_{ta} of the target and the contribution μ_{fl} of all flankers. The different congruency conditions in the flanker task were realized by assuming that μ_{fl} is positive for congruent flankers but negative for incongruent flankers. Under ideal conditions more attentional weight is presumably given to the target position than to the flanker positions. However, the exact distribution of the attentional weights depends on various factors (cf. B. A. Eriksen & Eriksen, 1974). In any case, in contrast to Gratton et al. (1992), but in accord with Cohen et al. (1992), we assume that some spatial filtering is already possible at stimulus onset.

Thus, in the present form, the DSTP model has seven free parameters: four rate parameters, two criteria, and a nondecisional parameter. The four rates are the component rates μ_{ta} and μ_{fl} for the target and flanker, respectively, the rate μ_{SS} for the late stimulus selection process (*SS*), and the rate μ_{RS2} for the response

selection process (*RS*2) in Phase 2. Because we assumed symmetric boundaries, criterion *A* represents the boundaries for the response selection process, whereas criterion *C* represents the boundaries for the late stimulus selection process. We also needed a nondecisional parameter t_{er} that represents the time used for stimulus encoding, response execution, and so forth. Finally, the diffusion coefficients for the processes also had to be specified. Because they are scaling parameters which are not identifiable, they were set to 0.01.

To evaluate the performance of our model, we wanted to fit not only the DSTP model to our data but also several alternative models. For the DSTP model an analytic version was available (see Appendix A), but this was not the case for most of the other models. Therefore, the performance of these models had to be simulated. Because the performance of models that have been fit to data by different procedures is difficult to compare, we decided to simulate all models.[1] They were programmed in C++ and fit with the PRAXIS algorithm (Brent, 1973; Gegenfurtner, 1992) to the quantiles of the cumulative distributions of correct responses and error responses. The alternative models and their performances are considered in a later section. In this and the next two main sections we merely consider the fits of the DSTP model to the data.

Following Ratcliff and Smith (2004), we estimated parameter values by minimizing the Wilks likelihood ratio chi-square, G^2, which approximates the chi-square statistics as sample sizes become large. With this goodness-of-fit measure both the distribution shapes and the response probabilities are taken into account simultaneously. Because for each spacing condition the distributions for congruent and incongruent stimuli were fit with one parameter set, we had seven free parameters for the DSTP model to fit 20 data

---

[1] A running program of the simulation version of the DSTP model can be found at http://www.psychologie.uni-konstanz.de/abteilungen/ kognitive-psychologie/research/projects/dstp-model/

points. This resulted in a total of 15 degrees of freedom (two distribution pairs, each with 11 degrees of freedom, minus seven free parameters of the model). However, like Ratcliff and Smith, we have used the G^2 statistic as a relative rather than an absolute measure of fit.

We simulated 8×10^5 trials per condition for each iteration. To prevent the obtained parameter estimates from representing a local minimum, we repeated the fit procedure several times with different sets of initial parameter values. In addition, we also fit the analytic version of our model to the data with the same fit algorithm as for the model simulations. This allowed us to assess the precision of our simulations. The goodness-of-fit measures for the two versions of the DSTP model are provided later in Table 7 (that for the analytic version is shown in parentheses). As can be seen, both measures are rather similar. Because this also holds for the estimated parameters, we report only the performance and the parameters obtained by the simulations.

For the standard diffusion model Ratcliff and colleagues have shown that the parameter estimates obtained by fitting the model to quantile-averaged data are similar to the averages of parameter estimates obtained by fitting the model to individual participant data (e.g., Ratcliff, Thapar, & McKoon, 2003; Thapar, Ratcliff, & McKoon, 2003). Because we did not know whether this result could be generalized to the DSTP model, we also compared the two estimation procedures. As can be seen in Table 1, the two sets of parameter values are rather similar. The values obtained for the averaged data are within half a standard deviation of the parameters obtained by averaging the values estimated for the individual participants. Therefore, in the following, we consider only the fit to the quantile-averaged data.

The performance of the DSTP model is represented by the lines in Figure 3. As can be seen, the fit is rather good. If we consider the parameter values, then we see that, as expected, the partial rate μ_{fl} for the flankers is substantially increased for the narrow condition, whereas the partial rate μ_{ta} for the target is similar between the spacing conditions (see Table 1). As a consequence, the overall rate for $RS1$ is even negative for incongruent stimuli in the narrow condition. The criterion is only slightly increased in the narrow condition. Concerning the late stimulus selection process, the rates are similar for the two spacing conditions. However, the criterion is increased under the narrow spacing. Moreover, the rate of the response selection process in Phase 2 was considerably higher for the narrow condition. This indicates a strong effect of late stimulus

selection on response selection, especially for the narrow spacing. Thus, overall, the parameter differences between the two spacing conditions are as expected.

In Table 2 the empirical and theoretical mean RTs and response probabilities are provided. As can be seen, the mean values are rather similar. Because our model allows one to calculate the relative contributions of $RS1$ and $RS2$ to performance, Table 2 also shows the probabilities p_{RS1}(A) and p_{RS1}(B) that a correct or incorrect response is already selected by $RS1$, respectively. As can be seen, for congruent stimuli $RS1$ selects more correct responses in the narrow condition than in the wide condition, whereas the opposite holds for incongruent stimuli. This results from the increased contribution of the flankers in the narrow condition. If we consider the error responses, then we see that almost all are produced by $RS1$. In the narrow condition, for instance, incongruent stimuli produced about 15.5% errors, of which 15.4% can be attributed to $RS1$.

All together, our model fit and analyses revealed that the DSTP model accounts rather well for the various aspects of the present data. The next experiment provides another example, where the efficiency of early spatial filtering was manipulated by varying spatial uncertainty and eccentricity.

Experiment 2: Spatial Uncertainty and Eccentricity of the Stimulus

In our second experiment we tried to manipulate the relative contributions of early and late stimulus selection in the flanker task by varying spatial uncertainty and eccentricity. For an effective spatial filtering it is important that the filter can be positioned optimally at the target location. Accordingly, spatial uncertainty about the exact location of target and flankers usually leads to an impaired performance (e.g., P. Brown & Fera, 1994; Goolkasian & Bojko, 2001; Miller, 1991; Paquet & Lortie, 1990). Thus, it is reasonable to assume that under spatial uncertainty a reliable performance can be achieved only if response selection is strongly supported by a late stimulus selection stage, where selection is based, for instance, on categorical spatial information such as "the middle letter."

To investigate whether late selection is indeed strongly involved under spatial uncertainty, we conducted an experiment in which the entire pattern of the target and the two flankers could occur at different positions on the screen. In one blocked condition (1-position-central)

Table 1

Parameter Values of the DSTP Model for the Stimulus Spacing Conditions in Experiment 1

	Parameter						
	Early selection			Late selection			
Condition	μ_{ta}	μ_{fl}	A	μ_{SS}	C	μ_{RS2}	t_{er}
Wide	.0404 (.0394)	±.0162 (±.0224)	.0681 (.0694)	.3657 (.3635)	.0706 (.0726)	±1.149 (±1.261)	.2155 (.2176)
Narrow	.0449 (.0372)	±.0651 (±.0663)	.0699 (.0717)	.3680 (.3717)	.0863 (.0874)	±1.575 (±1.524)	.2248 (.2255)

Note. The values in parentheses are the averages of the parameter estimates obtained by fitting the model to the individual data sets of the participants. DSTP = dual-stage two-phase; μ = rate or drift of a given diffusion process and condition; ta = target; fl = flanker; A = criterion for response selection; SS = stimulus selection process; C = criterion for stimulus selection; $RS2$ = response selection process in Phase 2; t_{er} = nondecisional parameter that represents time used for stimulus encoding, response execution, and so forth.

Table 2
Empirical Mean Data in Experiment 1 and Corresponding Values Produced by the DSTP Model

Condition	Congruent				Incongruent			
	$p(A)$	RT	$p_{RS1}(A)$	$p_{RS1}(B)$	$p(A)$	RT	$p_{RS1}(A)$	$p_{RS1}(B)$
Wide								
Data	.918	427			.899	435		
Model	.927	432	.161	.070	.902	436	.133	.096
Narrow								
Data	.932	444			.841	472		
Model	.945	449	.249	.054	.845	479	.113	.154

Note. The columns labeled with $p_{RS1}(A)$ contain the probabilities that a correct response is already selected in Phase 1 of response selection. The columns labeled with $p_{RS1}(B)$ provide the probabilities that an incorrect response is selected in Phase 1 of response selection. RT = response time, in milliseconds; DSTP = dual-stage two-phase.

there was no spatial uncertainty, because the stimuli always appeared at the center of the display. In another blocked condition (2-positions-lateral) the stimulus pattern could occur at either a left or at a right position. Although there was some spatial uncertainty, the two positions were chosen in such a way that the innermost flanker always appeared at the center of the display. Consequently, the five positions that could be occupied by items were uniquely linked to the target or to the flankers. Finally, there was a blocked condition (three positions) where the stimulus pattern either appeared at the center (3-positions-central) or at one of the two lateral positions (3-positions-lateral). This not only increased spatial uncertainty, but the middle three item positions were also no longer uniquely linked to the target or to flankers. With these features, the 3-positions condition had the largest spatial uncertainty. Thus, to assess the effect of uncertainty, we compared performance in the 1-position-central and 2-positions-lateral conditions with that in the 3-positions-central and 3-positions-lateral conditions, respectively. The details of the method and mean results can be found in Appendix B.

Obviously, the different conditions differed not only with respect to spatial uncertainty but also with respect to eccentricity, which should also affect stimulus selection. For a laterally presented stimulus pattern it can be expected that target selection is more difficult than for a pattern that appears at the center. Accordingly, the effects of eccentricity were assessed by comparing the performance between the central and lateral conditions.

The analysis of the mean performance supported our hypotheses that spatial uncertainty as well as eccentricity increased the difficulty of target selection (see Appendix B). As expected, the congruency effect in RT was significantly larger in the 3-positions conditions compared to the 1-position-central and 2-positions-lateral conditions. Moreover, the congruency effect was also larger for lateral than for central stimuli.

If we consider the distributions of the different conditions (see Figure 4), then it is obvious that the overall pattern of each condition is similar to that of the narrow condition in our first experiment. Additionally, the effects of eccentricity and uncertainty can be seen. With respect to eccentricity, the distributions of the correct RTs are generally shallower for the lateral positions than for the central position. Moreover, the functions for congruent and incongruent stimuli are further separated, which reflects the larger congruency effects. In the distributions of incorrect RTs the increased congruency effect for the lateral positions is reflected by an increased proportion of errors for incongruent stimuli. The increased congruency effects caused by

spatial uncertainty can be seen by the larger separation between the corresponding distributions in the 3-positions-central condition, compared to the 1-position-central condition. In the lateral conditions, this effect is relatively small. All together, though, our manipulations of selection difficulty had the intended effects. To see whether they are also reflected by the model parameters, the model was fit to our data.

Model Fit

The DSTP model was fit to the quantile-averaged distributions with the same procedure as before. The question was to what extent spatial uncertainty and eccentricity reduced the efficiency of early selection and whether this deficit was compensated for by an increased involvement of late selection. As can be seen in Figure 4, the model fit the data quite well. If we consider the obtained parameters for the 1-position-central condition (see Table 3), then we see that the partial rate for the flankers is smaller than that for the target. This is different from the comparable narrow condition in the previous experiment and indicates that in the present experimental context the participants allocated a larger portion of their attention to the target position than to the flanker positions. Accordingly, the rate for Phase 2 could also be smaller, compared to our first experiment.

This is also rather different from the situation where the stimulus pattern appeared at one of the lateral positions. As expected, in this case the attentional weight was relatively low for the target but high for the flankers. As a result, the overall rate for RS1 is negative for incongruent stimuli. This led to an increase of the response criterion and to a high rate for RS2. The same relation holds between the 3-positions-central and the 3-positions-lateral conditions. If we compare the 1-position-central condition with the same stimulus condition under spatial uncertainty (3-positions-central), then it is obvious that the distribution of attention was slightly changed in favor of the flankers, which caused the larger congruency effect. Obviously, for the lateral condition, spatial uncertainty had only a small effect (2-positions-lateral vs. 3-positions-lateral). Although the parameters indicate a stronger involvement of late selection in the 3-positions-lateral condition, the parameters for early selection are rather similar.

Thus, concerning the relative contributions of RS1 and RS2 to the overall performance, we have a similar picture as before (see Table 4). Under difficult selection conditions the overall rate for RS1 is rather small or even negative. Accordingly, only a few correct responses to incongruent stimuli are triggered by this
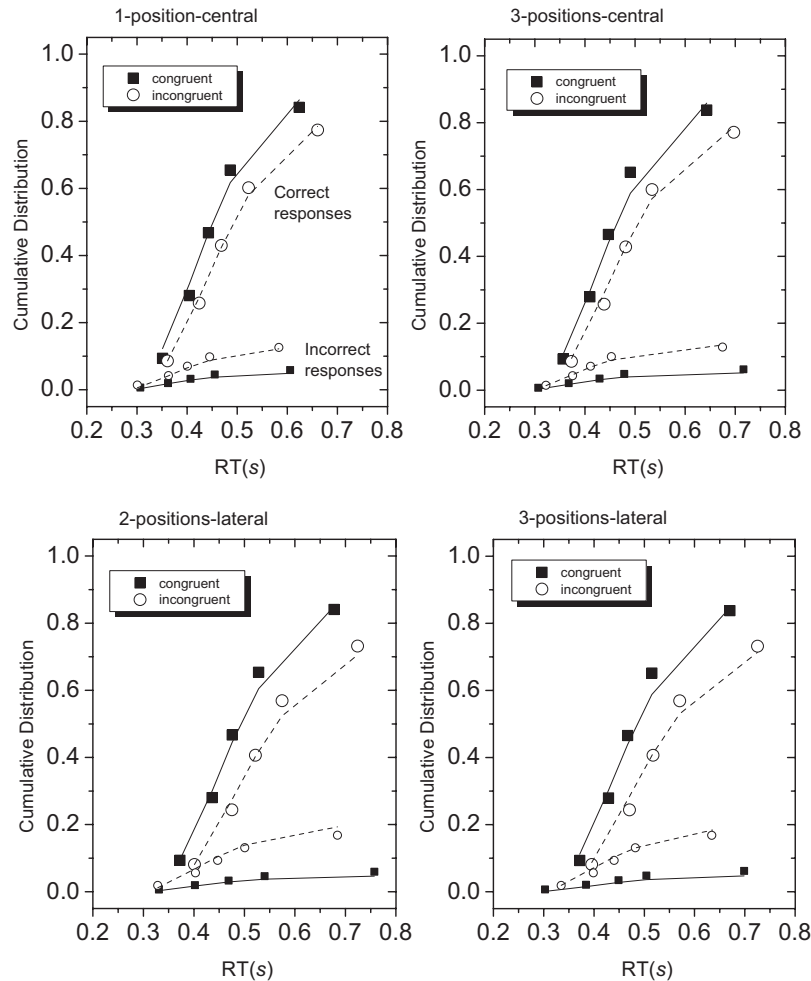
Figure 4. The panels show defective cumulative response time (RT) distributions (quantile averaged) separately for congruent and incongruent trials of the four conditions in our second experiment. The two upper distributions in each panel refer to correct responses, whereas the two lower distributions in each panel refer to error responses. Symbols represent experimental data and lines represent fit values of the dual-stage two-phase (DSTP) model.

process. In the 2-positions-lateral condition, for instance, only 9% of the correct responses were selected by $RS1$. In contrast, almost all incorrect responses are due to $RS1$. If we again consider the 2-positions-lateral condition, then of the 21% errors, 20.2% were caused by $RS1$. Thus, the small rate of $RS1$ reduces the probability that a correct response is already selected in Phase 1, but, at the same time, it also increases the probability that $RS2$ comes into play. Because response selection in Phase 2 is slow but reliable, this explains why the accuracy improves considerably with RT in this difficult condition, even though at some costs in RT.

All together, the results of the present experiment show that spatial uncertainty and eccentricity can have a substantial negative effect on performance. By fitting our DSTP model to the distributional data (see Figure 4) we were able to decipher how the attentional selection mechanisms adapt to the different stimulus conditions. It seems that, if a target always uniquely appears at a central location, then spatial attention can be effectively restricted to that location. This largely prevents the processing of the flankers. As a consequence, a response

can often be selected by the information provided by this early filtering process, even when the flankers are incongruent. However, if the target position is variable and shared with the flankers, and/or the target is presented at a lateral position, then spatial filtering is less efficient. For incongruent stimuli this leads to a reduced or even negative overall rate. Fortunately, the impaired spatial filtering can partly be compensated for by a strengthened late stimulus selection process, which keeps the performance at an acceptable level, at least for slow responses.

Experiment 3: Proportion of Congruent and Incongruent Stimuli

By analyzing and modeling the data of the last two experiments we have demonstrated that late selection can compensate deficits of early selection. It turned out that in conditions where early selection was not very selective, late selection was strengthened. Does this mean that any reduction of early selectivity always leads

Table 3
Parameter Estimates Obtained by Fitting the DSTP Model to Quantile-Averaged Data for the Different Conditions in Experiment 2

	Parameter						
	Early selection			Late selection			
Condition	μ_{ta}	μ_{fl}	A	μ_{SS}	C	μ_{RS2}	t_{er}
1-position-central	.0860	±.0562	.0666	.3240	.0820	±1.031	.2360
2-positions-lateral	.0352	±.0844	.0756	.2907	.0861	±1.252	.2434
3-positions-central	.0606	±.0565	.0706	.3053	.0775	±1.016	.2347
3-positions-lateral	.0364	±.0821	.0756	.2913	.0870	±1.253	.2349

Note. DSTP = dual-stage two-phase; μ = rate or drift of a given diffusion process and condition; ta = target; fl = flanker; A = criterion for response selection; SS = stimulus selection process; C = criterion for stimulus selection; $RS2$ = response selection process in Phase 2; t_{er} = nondecisional parameter that represents time used for stimulus encoding, response execution, and so forth.

to an enhanced late selection? Or are there also situations in which both selection mechanisms are adjusted in the same direction? As we have seen in the last experiments, the compensation of an early selection deficit is actually necessary only if the flankers carry incongruent information. With congruent flankers, low early selectivity even has an advantage, because it increases the overall rate of the response selection process. In this case a strong late selection process is of little help. Thus, in a situation where mainly congruent stimuli occur, we would expect a low early selectivity and, at the same time, only a low to moderate involvement of late selection. In contrast, when incongruent stimuli are frequent, then early selectivity should generally be high, and, if this is not sufficient for a reliable performance, late selection should be strong as well. These predictions were tested in the present experiment. While keeping the basic stimulus conditions constant, we tried to manipulate the selection strategy by varying the proportion of congruent trials.

For several conflict paradigms it has been shown that the proportion of congruent trials systematically affects performance (e.g., Gratton et al., 1992; Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979; Logan, Zbrodoff, & Williamson, 1984). These and related effects have been interpreted in the way that attention is generally less selective when stimuli are congruent in the majority of trials (see also Heitz & Engle, 2007; Mattler, 2006). Although this attentional control account has also been questioned (e.g., Mayr, Awh, & Laurey, 2003; Schmidt & Besner, 2008; Schmidt, Crump, Cheesman, & Besner, 2007), several studies strongly support it (e.g., Crump, Gong, & Milliken, 2006; Fernandez-Duque & Knight, 2008; Lehle & Hübner, 2008; Wendt, Kluwe, & Vietze, 2008). Thus, for our present objective we assumed that selectivity is determined by an attentional control strategy.

In our experiment we had one condition with 80% congruent trials and another condition with 20% congruent trials. The stimuli and basic procedure were similar to the 2-positions-lateral condition in the previous experiment (see Appendix B). We expected that the participants would use a specific selection strategy for each condition. If the stimuli are congruent on most of the trials, then stimulus processing should be less selective than when the stimuli are mostly incongruent. This strategic difference should be reflected by the flanker congruency effects. Indeed, in the 80%-congruent condition the flanker congruency effect in RT was significantly larger than in the 20%-congruent condition (see Appendix B). If we consider the RT distributions in Figure 5, then it is obvious that the congruency effect again increased with RT in

Table 4
Empirical Mean Data in Experiment 2 and Corresponding Values Produced by the DSTP Model

	Congruent				Incongruent			
Condition	$p(A)$	RT	$p_{RS1}(A)$	$p_{RS1}(B)$	$p(A)$	RT	$p_{RS1}(A)$	$p_{RS1}(B)$
1-position-central								
Data	.935	456			.860	484		
Model	.948	464	.319	.050	.865	493	.193	.127
2-positions-lateral								
Data	.934	493			.813	536		
Model	.952	512	.286	.047	.790	547	.090	.202
3-positions-central								
Data	.931	460			.857	497		
Model	.947	479	.282	.051	.857	505	.142	.133
3-positions-lateral								
Data	.931	484			.813	530		
Model	.950	505	.291	.0485	.793	540	.098	.197

Note. The columns labeled with $p_{RS1}(A)$ contain the probabilities that a correct response is already selected in Phase 1 of response selection. The columns labeled with $p_{RS1}(B)$ provide the probabilities that an incorrect response is selected in Phase 1 of response selection. RT = response time, in milliseconds; DSTP = dual-stage two-phase.
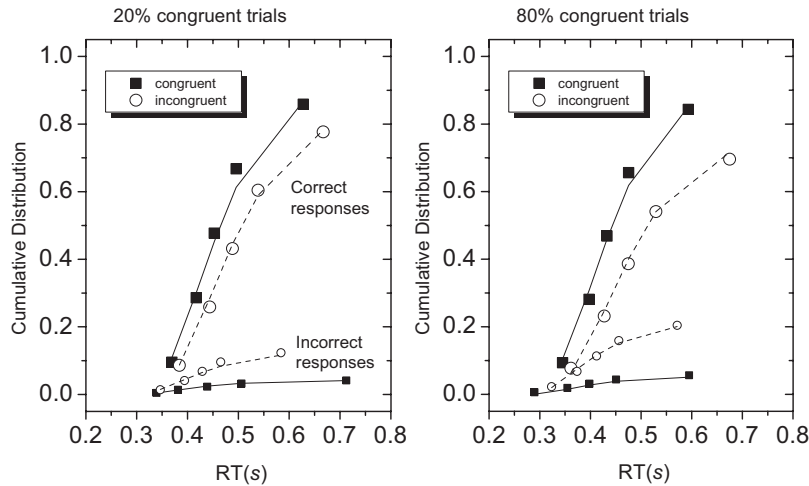
Figure 5. The two panels show defective cumulative response time (RT) distributions (quantile averaged) separately for congruent and incongruent trials of the two conditions in our third experiment. The two upper distributions in each panel refer to correct responses, whereas the two lower distributions in each panel refer to error responses. Symbols represent experimental data and lines represent fit values of the dual-stage two-phase (DSTP) model.

both conditions, but to a larger extent in the 80%-congruent condition. The congruency effect also differed substantially for error responses between the conditions, as can be seen by considering the corresponding distributions (see Figure 5).

The reduced selectivity in the 80%-congruent condition sped up responding, relative to the 20%-congruent condition. However, if we consider the distributions in Figure 5, then we see that in the 80%-congruency condition many of the fast responses to incongruent stimuli were errors. Thus, whereas the applied attentional strategy is reflected in the error rates by costs for fast responses, in RT it is mainly reflected by costs for slow responses. The interesting question was whether the reduced early selectivity in the 80%-congruent condition was this time not compensated for by a strengthened late selection process, compared to the 20%-congruent condition. To see whether this was the case, we fit the DSTP model to the data.

Model Fit

In Table 5 the parameter values estimated by the same procedure as before are shown. As expected, the values indicate that the selectivity of the spatial filter is relatively low for the 80%-

congruent condition (see Table 5), which leads to a negative overall rate for incongruent stimuli. Because the criterion for the response selection process is also relatively low, a large proportion (42.8%; see Table 6) of the correct responses to congruent stimuli are already selected by $RS1$. This explains the good performances for congruent stimuli in the 80%-congruent condition. However, as already mentioned, this strategy produces costs for incongruent stimuli. The negative overall rate combined with the relatively low criterion, produces many fast errors, as reflected by the corresponding distribution. In case no response is selected in Phase 1, $RS2$ starts relatively late. Because most of the correct responses to incongruent stimuli are selected by $RS2$, the congruency effect in RT is increased for slow responses.

If we consider the parameter values for the 20%-congruent condition, then we see that the selectivity of the spatial filter is increased (see Table 5) relative to that in the 80%-congruent condition. As a consequence, the overall rate for incongruent stimuli is small, but positive. Furthermore, the criterion for the response selection process is slightly increased. Although these parameter values are helpful for attaining an acceptable perfor-

Table 5

Parameter Estimates Obtained by Fitting the DSTP Model to Quantile-Averaged Data for the Different Conditions in Experiment 3

	Parameter						
	Early selection			Late selection			
Condition	μ_{ta}	μ_{fl}	A	μ_{SS}	C	μ_{RS2}	t_{er}
20% congruent	.0846	±.0656	.0691	.3351	.0856	±1.520	.2573
80% congruent	.0706	±.0928	.0616	.3181	.0820	±1.342	.2522

Note. DSTP = dual-stage two-phase; μ = rate or drift of a given diffusion process and condition; ta = target; fl = flanker; A = criterion for response selection; SS = stimulus selection process; C = criterion for stimulus selection; $RS2$ = response selection process in Phase 2; t_{er} = nondecisional parameter that represents time used for stimulus encoding, response execution, and so forth.

Table 6

Empirical Mean Data in Experiment 3 and Corresponding Values Produced by the DSTP Model

	Congruent				Incongruent			
Condition	p(A)	RT	p_{RS1}(A)	p_{RS1}(B)	p(A)	RT	p_{RS1}(A)	p_{RS1}(B)
20% congruent								
Data	.954	472			.863	504		
Model	.958	480	.312	.041	.867	509	.138	.130
80% congruent								
Data	.937	449			.773	492		
Model	.946	456	.428	.053	.775	493	.165	.224

Note. The columns labeled with p_{RS1}(A) contain the probabilities that a correct response is already selected in Phase 1 of response selection. The columns labeled with p_{RS1}(B) provide the probabilities that an incorrect response is selected in Phase 1 of response selection. RT = response time, in milliseconds; DSTP = dual-stage two-phase.

mance on incongruent trials, which make up the majority, the achieved early selectivity is obviously not sufficient to meet the demands in such experimental conditions. Therefore, late selection is strengthened as well by increasing the rate for the stimulus selection process (and for *RS*2), by increasing the rate of the stimulus selection process, and by slightly increasing the criterion (*C*). Consequently, *RS*1 selects a correct response on relatively few congruent trials (31.2%), which is also responsible for the increased mean RT. For incongruent stimuli, *RS*1 even selects a correct response on only 13.8% of the trials. Thus, by strengthening both early and late stimulus selection, the selection of a response by *RS*1 is largely avoided, which guarantees relatively high accuracy.

All together, the estimated parameter values for the DSTP model for the different ratio conditions show that late selection is not automatically strengthened when early selectivity is reduced. Under certain conditions, for example, when flanker coprocessing is advantageous and the risk of committing an error is low, both early selection and late selection are reduced. On the other hand, if the conditions for achieving a reliable performance are relatively difficult, both selection mechanisms are strengthened, if possible.

Comparison With Alternative Models

The application of our DSTP model to empirical data demonstrates that it can account well for the details of various attentional effects. However, because it has a relatively complex structure that includes several mechanisms of attentional selection, one might ask whether this complexity is really necessary. Perhaps an advanced single-stage selection model could also account for the considered distributional data. And even if this is not the case, there might be alternative models of similar complexity as the DSTP model that could explain the distributional data. Although these questions cannot exhaustively be answered in this article, we provide at least partial answers. First we examine the performance of a single-stage model. Then we examine models with a continuously increasing drift. Finally, we consider two variants of the DSTP model.

Comparing different models is not an easy task, because they usually differ in various aspects, especially if they are expressed in different formal frameworks. Therefore, to make the comparison easier, we implemented all considered models in the framework of diffusion processes and fit them to our distributional data in the

same way as the DSTP model. The performance of the models was assessed by means of the Bayesian information criterion (BIC) model-selection statistic (G. Schwarz, 1978), which takes the number of free model parameters into account. There are also other statistics (cf. Zucchini, 2000), which, however, would lead to the same conclusions in the present case. We preferred the BIC because it can be derived from the G^2 statistic (cf. Ratcliff & Smith, 2004) that we used for data fitting. The G^2 and BIC statistics for all models and fits are listed in Table 7. The best model is the one with the smallest BIC.

Single-Stage Model

That single-stage models can account for the mean effects of flanker congruency has already been shown (e.g., Logan, 1996). Here, we examine to what extent such models can also account for distributional data. It can be expected that single-stage models have problems obtaining an acceptable mean accuracy for incongruent stimuli in difficult selection conditions and producing a large error rate for fast responses at the same time. To see how a single-stage model performs, we combined our early selection mechanism with a stationary diffusion process for response selection. However, a simple diffusion process can principally not account for faster error responses than correct responses, because the corresponding distribution functions are identical. Therefore, we assumed a random starting value for the diffusion process, which, on average, leads to faster errors than correct responses (cf. Ratcliff, 1981). More specifically, we assumed that $X(0)$ is a uniformly distributed random variable with mean zero and range s_z. Thus, all together we had five free model parameters: the component rates for target and flankers, the response criterion, the range of the starting value, and the nondecisional time.

Fitting the single-stage model to the distributional data of all experimental conditions revealed that, as expected, its performance was relatively poor (see Table 7). To show an example, we plotted the performance of this model for the narrow condition in our first experiment (see Figure 6). As can be seen, the model substantially overestimates the accuracy for responses to incongruent stimuli. Also, the fits to the distributions of the correct responses deviate systematically. The proportion of fast correct responses is overestimated, whereas that for slow correct responses is underestimated. If we consider the obtained parameter estimates (narrow condition: $\mu_{ta} = .2455$, $\mu_{fl} = \pm.0344$, $A = .0526$, $s_z =$

Table 7
Fit Statistics of the Considered Models for the Different Experimental Conditions

Condition and model	G^2	df	BIC
Wide			
DSTP	12.4 (12.3)	15	59.9 (59.7)
No priming	12.6	15	60.0
Race	16.4	15	63.9
Nonlinear increase	26.6	13	81.2
Linear increase	31.6	15	79.1
Single stage	52.9	17	86.8
Narrow			
DSTP	15.7 (15.8)	15	63.2 (63.2)
No priming	16.3	15	63.8
Race	18.0	15	65.4
Nonlinear increase	42.3	13	103
Linear increase	44.0	15	91.5
Single stage	68.6	17	103
1-position-central			
DSTP	7.86 (7.33)	15	50.3 (49.7)
No priming	7.37	15	49.8
Race	7.87	15	50.3
Nonlinear increase	18.8	13	73.3
Linear increase	25.1	15	67.5
Single stage	33.6	17	63.9
2-positions-lateral			
DSTP	10.0 (10.3)	15	52.5 (52.7)
No priming	10.4	15	52.9
Race	11.5	15	53.9
Nonlinear increase	25.5	13	80.0
Linear increase	49.6	15	92.0
Single stage	39.0	17	69.3
3-positions-central			
DSTP	8.18 (7.74)	15	47.7 (47.3)
No priming	8.19	15	47.7
Race	8.77	15	48.3
Nonlinear increase	19.2	13	70.0
Linear increase	22.3	15	62.3
Single stage	21.2	17	49.4
3-positions-lateral			
DSTP	17.7 (16.8)	15	62.2 (61.3)
No priming	18.3	15	62.8
Race	19.1	15	63.6
Nonlinear increase	40.7	13	97.9
Linear increase	64.8	15	109
Single stage	66.8	17	98.6
20%-congruent			
DSTP	12.6 (12.1)	15	61.6 (61.0)
No priming	13.0	15	61.9
Race	13.5	15	62.4
Nonlinear increase	25.7	13	89.7
Linear increase	54.6	17	104
Single stage	60.4	17	95.3
80%-congruent			
DSTP	25.3 (24.8)	15	74.3 (73.9)
No priming	24.7	15	73.8
Race	24.7	15	73.8
Nonlinear increase	51.1	13	114
Linear increase	95	15	144
Single stage	102	17	137

Note. The G^2 and Bayesian information criterion (BIC) values are based on the performance obtained by computer simulations of the models. The values in parentheses for the dual-stage two-phase (DSTP) model represent the corresponding values obtained with the analytic version (see Appendix A) of the model.

.0566, t_{er} = .2848; wide condition: μ_{ta} = .3049, μ_{fl} = ±.0105, A = .0539, s_z = .0594, t_{er} = .274548), then we see that the component rates and the criteria vary in the expected direction between the two spacing conditions, but only moderately. For instance, even for the narrow condition the overall rate for incongruent stimuli differs by only a relatively small amount from that for congruent stimuli. This explains the overestimated accuracy for the former condition. Thus, the single-stage model cannot account for results where performance is poor for fast responses but good for slow responses.

For comparison, we also fit the Ratcliff diffusion model to the data, that is, a diffusion model where not only the starting value but also the rate and t_{er} vary randomly across trials. It turned out that these additional parameters did not improve the performance. Thus, it seems that a standard single-stage model is not appropriate for explaining our data.

Continuously Increasing Drift Model

The poor performance of the single-stage model indicates that at least for conflict paradigms it is necessary to assume some form of increase in selectivity during stimulus processing. In the DSTP model selectivity improves abruptly in one step at some point in time. Because this point was interpreted as the time when the output of a categorical late selection mechanism is available, the success of our model can be considered support for the idea that selection can occur either at an early or at a late stage. This idea is also in accord with Gratton et al.'s (1992) dual-phase idea, that is, with the assumption that there are two phases of different selectivity. If we consider the various formal approaches to account for Gratton et al.'s (1988) distributional data, though, then we see that all of them assume a gradually increasing selectivity. One reason for this characteristic might be that neural networks were mostly used for modeling. Cohen et al. (1992; see also Servan-Schreiber, Bruno, Carter, & Cohen, 1998), for instance, used such a framework. Although the performance of their model was qualitatively in accord with the data, there were relatively large and systematic quantitative deviations. For instance, the accuracy for the congruent condition was systematically overestimated. Some points of the fit even deviate by more than 10% (see Figure 9 in Cohen et al., 1992). Furthermore, the details of this approach remain unclear, for example, how many free parameters were used and exactly how selectivity changes over time.

Recently, Liu et al. (2008) also modeled Gratton et al.'s (1988) data using a neural network model. However, because their model involves somewhat arbitrary assumptions on architecture and parameters, the authors also approximated its behavior with a diffusion model with a linear, quadratic, or exponential drift rate. It turned out that the diffusion model with the exponential drift rate best accounted for the data, although there were still systematic deviations. Here, we apply similar types of models to fit our data. We assumed that response selection proceeds by a diffusion process whose rate can change gradually with time. In our first approach, we further assumed that the contribution of the flankers continuously vanishes with time, which corresponds to the idea of a dynamic spotlight (Heitz & Engle, 2007) whose diameter shrinks gradually until it is focused on the target.

For incongruent stimuli, where the initial effect of the flankers is negative, these assumptions lead to a gradual increase of the
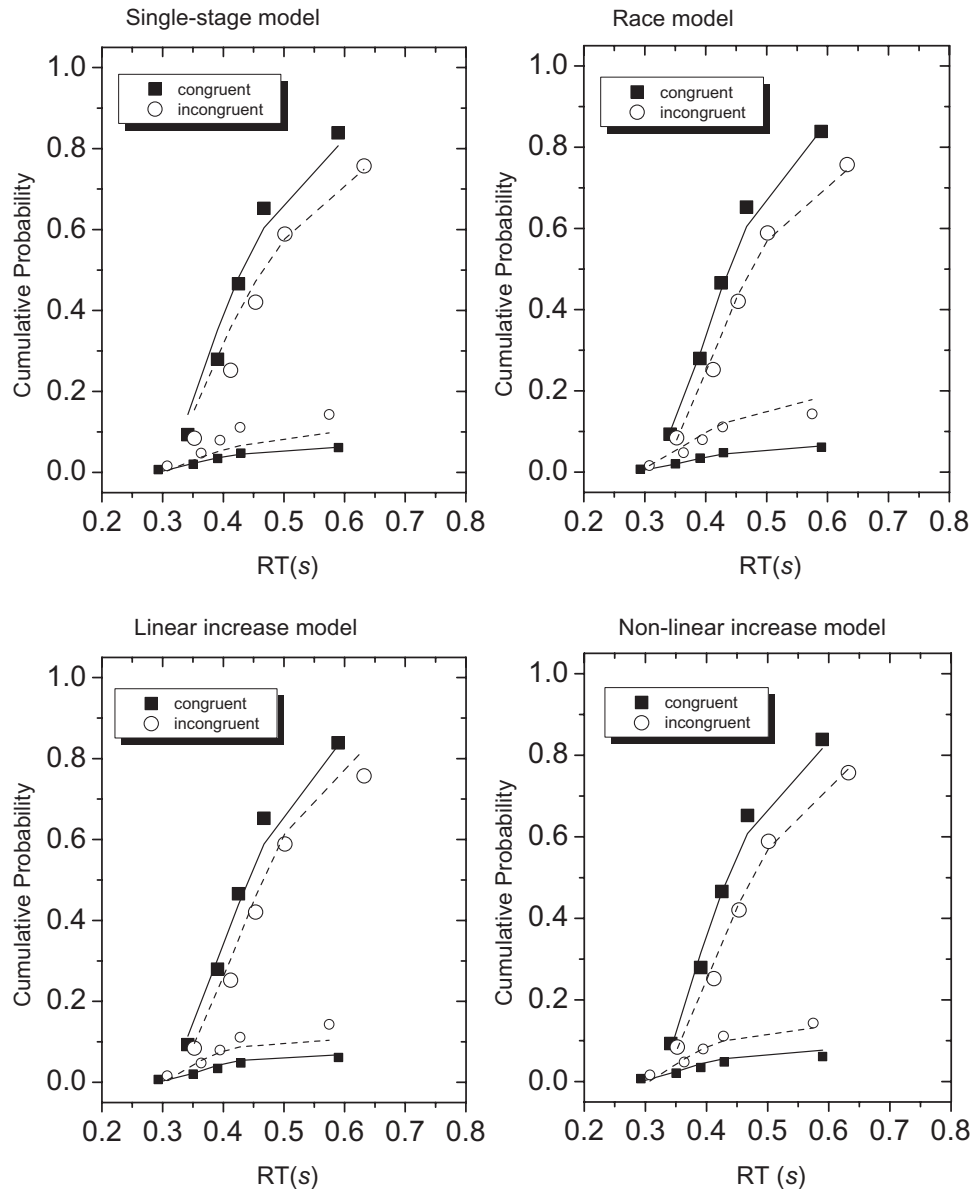
Figure 6. The panels show examples of the performance of the alternative models. The symbols in each panel represent the data of the narrow spacing condition in our first experiment. The lines in each of the four panels represent the fit values of one of the alternative models as identified above each panel.

overall rate until the spotlight is restricted to the target. For congruent stimuli, though, where the initial effect of the flankers is positive, such a dynamic would imply that the overall rate decreases, which seems unreasonable. To avoid such a consequence, it is usually assumed that the density of the attention resources increases with a decreasing diameter of the spotlight (C. W. Eriksen & St. James, 1986; Heitz & Engle, 2007). For the present framework this implies that the attentional weight for the target increases with a shrinking spotlight. If the weight for the target increases to the same extent as that for the flankers decreases, then the overall rate for congruent stimuli remains constant. Accordingly, for congruent stimuli we assumed that the overall rate of the response selection process has a constant value μ_x, whereas the

rate for incongruent stimuli starts with a relatively small value and then increases until it also reaches the value μ_x.

But exactly how should the rate increase? We first considered a model with a linearly changing rate (cf. Spieler et al., 2000). For incongruent stimuli the overall rate was set to $\mu(t) = \mu_x - \mu_y(t)$ for $t \geq 0$, where $\mu_y(t)$ decreases linearly with time, that is, $\mu_y(t) = \mu_{y0} - t\delta_y$, until it reaches zero. The parameter δ_y represents the decrement per time unit. However, the performance of this model was rather poor. Therefore, we additionally assumed that the overall rate further increases even after the spotlight has excluded the flankers, that is, after $\mu_y(t)$ has reached zero. Specifically, we assumed that the common rate μ_x increases from that point on by $t\delta_x$ up to a maximum rate μ_{max}. Including the boundary and

nondecisional time parameter t_{er}, this dynamic spotlight model has seven free parameters.

The result of the model fit is also shown in Table 7. As can be seen, compared to the simple single-stage model the fit has improved for most conditions. If we consider the performance for the narrow condition in our first experiment (see Figure 6), then it is obvious that in this case the fit mainly improved for the distributions of correct RTs. The fit to the distribution of the error RTs remains poor. The parameter estimates for this condition are as follows: $\mu_x = .1829$, $\delta_x = .0020$, $\mu_{max} = .3660$, $\mu_{y0} = .1911$, $\delta_y = .0024$, $A = .0501$, and $t_{er} = .2709$. For the wide condition the parameter estimates are as follows: $\mu_x = .1741$, $\delta x = .0016$, $\mu_{max} = .3622$, $\mu_{y0} = .1929$, $\delta_y = .0037$, $A = .0536$, and $t_{er} = .2402$. These values are difficult to interpret. For instance, the starting value μ_{y0} of the flankers is larger for the wide than for the narrow condition, and an even more pronounced relation holds for δ_y, which is not very plausible. Taken together, we can conclude that although the dynamic spotlight led to improved fits compared to the single-stage model, it remains rather poor in absolute terms. This demonstrates that the simple assumption of a linearly increasing selectivity, even if there are different periods with different slopes, is not appropriate for satisfactorily explaining our distributional data.

One might argue that the assumption of a stepwise linearly increasing rate and the other constraints are too restrictive and that a nonlinear increase with fewer constraints might be more appropriate. To see whether this is the case, we also considered a model with a nonlinearly increasing rate. But how should the rate increase? As mentioned, Liu et al. (2008) applied a quadratic and an exponential drift rate, but with moderate success. A proper function might be obtained by examining how the mean rate (across trials) of the DSTP model changes with time (under the condition that no response has been selected yet). It turned out that this change can almost perfectly be described by a logistic growth function of the form $\mu(t) = (\mu_0 - \mu_{max})/[1 + (t/t_c)^a] + \mu_{max}$. In this equation the parameters μ_0 and μ_{max} are the starting and maximal value of the rate, respectively. The parameter t_c is the point of inflection, and the exponent a determines how fast the mean rate grows. We used this type of nonlinear function to model the growth in rate with time. Moreover, except for μ_{max}, we used individual parameter sets of this function for the two congruency conditions. Thus, including the common response criterion A and t_{er} we had nine parameters. As can be seen in Table 7, compared to the piecewise linear model the fit generally improved.

In Figure 6 the fit to the narrow condition in our first experiment is shown. The values of the common parameter values are $\mu_{max} = .2911$, $A = .0474$, and $t_{er} = .2842$. The values of the specific parameters for the congruent condition are $\mu_0 = .0010$, $t_c = .0232$, $a = .6006$, and those for the incongruent condition are $\mu_0 = .0002$, $t_c = .0548$, $a = .2246$. As one would expect, the starting rate (μ_0) is larger for the congruent condition than the incongruent one. Also the increase of the rate occurs earlier for the former condition and is faster, as indicated by the values of t_c and a.

If, for comparison, we consider the wide condition, then the values of the common parameters are $\mu_{max} = .2903$, $A = .0495$, and $t_{er} = .2586$. The values of the other parameters for the congruent condition are $\mu_0 = .0003$, $t_c = .0297$, $a = .4914$, and those for the incongruent condition are $\mu_0 = .0018$, $t_c = .0400$, $a = .2264$. Concerning the values of the parameters t_c and a, the

difference between the two congruency conditions is smaller compared with the narrow condition, which is plausible. The starting rate, however, is larger in the incongruent condition than in the congruent condition, which is implausible.

Nevertheless, the nonlinear model fit the data generally better than the piecewise linear model. However, due to the costs produced by the increased number of parameters, the BIC is smaller in only half of the cases (see Table 7). Moreover, the goodness of fit of the nonlinear model is still far away from that of the DSTP model. Clearly, we cannot definitely exclude that there is a nonlinear model that fits the data much better. However, it is hard to tell whether such a model exists, and even if there were such a model, it would be difficult to find and presumably very special. Thus, taken together, our considerations suggest that it is rather unlikely that a model with a gradually increasing selectivity can generally account well for distributional flanker task data.

Variants of the DSTP Model

The considerations so far suggest that the two discrete stages of stimulus selection and the corresponding two phases of response selection are the crucial characteristics of the DSTP model that are responsible for its successful performance. Additionally, however, we made a number of further specific assumptions. The question is how important these assumptions are. For instance, we assumed that there is only one response selection process that changes its rate after a stimulus has been selected. Alternatively, though, one could assume that a second (controlled) response selection process starts (with a starting value of zero) when a stimulus is selected and that the first (automatic) response selection process continues unchanged.

Such an architecture would be a realization of a specific variant of the dual-route or dual-process idea. As mentioned in the introduction, one version of this idea, which is structurally rather similar to the DSTP model, assumes that an automatic priming process and a controlled rule-based process affect the same response selection mechanism (e.g., De Jong et al., 1994; Hommel, 1998; Hunt & Lansman, 1986; Ridderinkhof, 1997). Another version, however, assumes that the automatic and controlled routes are independent and that each of the corresponding processes can select a response (cf. Logan, 1988; Nosofsky & Palmeri, 1997; Palmeri, 1997). Thus, the dual processes compete in a race, and the process that reaches the response criterion first determines the response. This idea can easily be implemented as a variant of the DSTP model. Instead of that $RS2$ takes over from $RS1$, one would simple have to assume that $RS2$ runs in parallel with $RS1$ after a stimulus has been selected until one of the processes selects a response. Such a race model has the same number of parameters as the DSTP model.

The performance statistics of this race model are also given in Table 7. As can be seen, in some cases the performance is worse than that of the DSTP model, in others it is rather similar, and in one case it is even better. As for the other models, the fit to the narrow condition in our first experiment is shown as an example (see Figure 6). The obtained parameter values for the narrow condition are as follows: $\mu_{ta} = .0452$, $\mu_{fl} = \pm.0647$, $A = .0694$, $\mu_{SS} = .3681$, $C = .0864$, $\mu_{RS2} = \pm1.5745$, and $t_{er} = .2247$; those for the wide condition are as follows: $\mu_{ta} = .0442$, $\mu_{fl} = \pm.0172$, $A = .0673$, $\mu_{SS} = .3777$, $C = .0724$, $\mu_{RS2} = \pm1.1606$, and $t_{er} =$

.2135. Obviously, the values are similar to those for the DSTP model. For instance, the overall rate for the narrow condition is negative and, compared to the wide condition, the rate for $RS2$ is larger. If we consider Figure 6, then it is obvious that the fit is slightly worse than that of the DSTP model. This is also reflected by an overestimation of the error rate in the incongruent condition (20% instead of 16%) and by an underestimation of the flanker congruency effect (15 ms instead of 30 ms).

Nevertheless, all together the performance of the race model is rather similar to that of the DSTP model. However, this is not surprising, if one takes into account that, according to the superposition principle (Diederich & Colonius, 1991; W. Schwarz, 1994), two parallel diffusion processes are equivalent to a single diffusion process. For this equivalence it is usually assumed that both processes start from the same value. In the present case, though, we have the situation that, considered from stimulus selection onward, $RS2$ starts from zero, but $RS1$ continues from a random state, that is, from the end state of the first phase. This fact is presumably responsible for the small difference in performance between the race and the DSTP model that was observed in our analysis.

These considerations raise the question of whether it is important that $RS2$ starts from the end state of $RS1$. This feature was included in the DSTP model to reflect the priming of the state of evidence by mechanisms driven by early information, which is a crucial characteristic of the continuous flow idea (C. W. Eriksen & Schultz, 1979). To examine whether this feature is important in the present case, we also simulated a variant of the DSTP model, where $RS2$ again starts at the time when $RS1$ ends, but this time always with a starting value of zero. If we consider the performance statistic of this *no-priming* model in Table 7, then we see that the difference to the DSTP model is rather small. In two cases the performance is even better. Thus, for fitting the present data the assumption of response priming is not generally important. However, our results suggest that response priming becomes more important the more late selection is involved in response selection.

General Discussion

Various results observed during the last decades indicate that the mechanisms of selective attention are rather flexible and adaptive. Depending on the circumstances and the specific task, they select stimulus information at an early and/or late stage of processing (cf. Kinchla, 1992; Pashler et al., 2001). Furthermore, results from conflict paradigms such as the flanker task suggest that information provided by the early and the late stages contributes to response selection in a dynamical manner. This can be concluded from distributional data (e.g., Gratton et al., 1992). They show that response-incompatible information mainly affects the accuracy of fast responses, which indicates that there are two phases of selectivity. In the first phase selectivity is poor, which produces fast but unreliable responses. In the second phase, however, selectivity is high, which leads to slow but reliable responses. Thus, it seems that selectivity improves in the course of stimulus processing.

Up to now, however, the flexibility and dynamics of selective attention have not been appropriately reflected by models of selective attention. Therefore, the aim of the present article was to introduce a dual-stage two-phase (DSTP) model that takes these properties into account. The basic idea is that there are two stages

of stimulus selection and that their output leads to two corresponding phases of response selection, which are represented by a nonstationary diffusion process. In the first phase after stimulus onset the rate of this process is determined by the output of the early stage, where stimulus information is selected by means of perceptual filters. During the second phase the rate of response selection is determined by the output of a late stimulus selection stage. Late stimulus selection is also represented by a diffusion process that runs in parallel to response selection and selects specific *categorical* information from the display. Because it is assumed that perceptual filtering is less effective than categorical filtering, there are two phases of different selectivity. As the late stimulus selection process and the response selection process compete in a race, it is possible that a response is already selected during the first phase, which usually leads to fast but unreliable responses, or during the second phase, which produces slow but reliable responses.

The basic idea that stimulus processing consists of discrete phases differing with respect to processing rate has also been used in other models (e.g., Busemeyer & Diederich, 2002; Diederich, 1997; Lamberts, 2000). In the model of multidimensional stimulus categorization of Lamberts and colleagues (Lamberts, 2000; for a recent overview, see Lamberts & Kent, 2008), for instance, performance depends on the similarity of stimulus features sampled from the current stimulus to those from exemplars stored in memory (see also Nosofsky & Palmeri, 1997). Stimulus features are sampled in a stepwise manner, and therefore, similarity changes in discrete steps. Although this resembles the change of processing rate across phases in our model, there are also a number of differences. Most important, in their model, the number of processed stimulus features is increasing in the course of stimulus processing. In contrast, our model implies that the number of stimulus elements affecting response selection is reduced in the course of processing. A similar reasoning also applies to multiattribute decision field theory (Busemeyer & Diederich, 2002; Diederich, 1997). Nevertheless, the similarities between the models demonstrate the general applicability of the idea of a nonstationary process of response selection.

Although the DSTP model is rather abstract and can potentially be applied to various paradigms, in the present study we restricted our considerations to selective visual attention in the flanker task (B. A. Eriksen & Eriksen, 1974). We reported a series of three experiments in which we manipulated the relative contributions of early and late selection to performance. In the three experiments this was achieved by varying the spacing between target and flankers, the spatial uncertainty and eccentricity, and the relative proportion of congruent and incongruent stimuli, respectively. Fitting our model to the obtained distributional data should reveal whether the estimated parameter values vary in the predicted direction. In the first two experiments it turned out that, as expected, a stronger involvement of the late selection process compensated for early selection deficits in difficult selection conditions. In our third experiment, though, we could show that such compensation is not obligatory. In the two experimental conditions the strength of early and late selection strategically changed in the same direction to improve overall performance.

All together, our results and analyses demonstrate that early and late selection mechanisms work hand in hand in a rather flexible way, depending on the demands of the task. That selectivity also

depends on the experimental context can be seen by the fact that, although the narrow condition in our first experiment was practically identical to the 1-position condition in our second experiment, the corresponding performances differed, especially with respect to the error RT distributions. This demonstrates that the experience with the specific conditions in an experiment can influence the stimulus selection strategy.

Our investigation also extends the view of which selection mechanism can be controlled and how this is achieved to adapt to task demands. It seems that individuals can not only control early selection by tuning their perceptual filters (e.g., B. A. Eriksen & Eriksen, 1974; C. W. Eriksen & St. James, 1986) but also that they can regulate the strength of late selection. According to our model, the latter can be achieved indirectly by regulating early selection and by adjusting the response criterion. That is, increasing the criterion might not only improve the accuracy of responses determined by the output of early selection processes, but the resulting prolonged processing may also increase the probability that late stimulus selection comes into play and improves response selection. However, our results also suggest that late selection can be controlled directly to some extent.

Possible Alternative Accounts

The DSTP model was mainly developed to overcome the limits of single-stage models of selection. To demonstrate that these limits really exist, we also fit a single-stage model to our distributional data and found that, as expected, the fit was rather poor. It remains poor even when a gradually increasing selectivity is assumed. Such models have especially been proposed in connection with the zoom-lens metaphor. A specific version is the dynamic spotlight of attention proposed by Heitz et al. (2007). These authors assumed that at the beginning of stimulus processing the spotlight always encompasses all items in the display and then gradually shrinks until it is focused on the target.

A similar idea has also been realized as neural network models (Cohen et al., 1992; Servan-Schreiber et al., 1998) to account for the Gratton et al. (1988) data (see also Yu et al., 2009). To examine the effects of a gradually increasing selectivity, we first considered a model with a piecewise linearly increasing rate and then a model with a nonlinearly increasing rate. Although these models fit the data better than the single-stage model, the fit was still worse than that of the DSTP model. Thus, taken together, our comparison suggests that models with a gradually increasing selectivity are not a compelling alternative to our DSTP model. Even if one could find an exemplar that fits the data quite well, the obtained solution would presumably be rather specific.

Thus, it seems that the strength of the DSTP model lies in its two discrete stages of selectivity and in the fact that the transition from low selectivity to high selectivity occurs randomly at some point in time. With these properties the model is highly flexible. That the details of the second phase are less important was demonstrated by considering two variants of our DSTP model. In one version it was assumed that the second phase is characterized by a race between the continuing response selection processes from the first phase and an additional response selection process. The additional process should represent the effect of late stimulus selection and increase the selectivity. This model, which is similar to a certain class of dual-route accounts (cf. Logan, 1988; Nosofsky & Palm-

eri, 1997; Palmeri, 1997), also fit the data rather well. The same holds for a version without priming, that is, a version where we assumed that response selection in the second phase always starts from a state of zero instead from the end state of Phase 1.

Even though the considered variants of the DSTP model also fit the data rather well, overall, the DSTP model turned out to be the most successful version. Moreover, although the present study does not provide a strong test between the original and its variants, the DSTP model is not only more simple but also more plausible. The assumption of a single response selection process that changes its rate with the state of stimulus selectivity is conceptually simpler than a race between two response selection processes. Furthermore, in view of the abundant results collected with the flanker task it is rather unlikely that there is no response priming. Besides behavioral data, electrophysiological results, especially lateralized readiness potentials, also support the existence of such a priming mechanism (cf. Gratton et al., 1988; Leuthold, Sommer, & Ulrich, 1996).

Related Models From the Stroop-Task Literature

From our overview in the introduction it became obvious that most models of the flanker task take mainly early selection mechanisms into account and largely neglect late selection. However, late selection mechanisms have often been used to explain performance in the Stroop task (MacLeod, 1991; Stroop, 1935). Therefore, it might be informative to see how early and late selection and their interaction have been conceptualized in this area.

The Stroop task requires participants to name the color of a word while ignoring the word's meaning, which also refers to a color. The Stroop effect denotes the finding that color naming performance is strongly impaired when the word's meaning is incongruent to the word's color (e.g., the word *red* presented in green) compared to when meaning and color are congruent (e.g., the word *green* presented in green). Compared to the flanker task, the relevant stimulus dimension in the Stroop task cannot be selected effectively by spatial or other perceptual filters. Therefore, most models assume that late selection is applied to accomplish the task. In the following, we consider some of these models and how they relate to our DSTP model.

Most recent models for the Stroop task were implemented within a pure connectionist framework (Cohen, Dunbar, & McClelland, 1990; Cohen & Huston, 1994; but see Melara & Algom, 2003; Phaf, Van der Heijden, & Hudson, 1990). In these models, activation spreads from stimulus units via units representing semantic categories to response units. A response is elicited when the accumulated evidence exceeds a response criterion (Cohen et al., 1990; Cohen & Huston, 1994) or the network settles in an attractor state associated with a response (Phaf et al., 1990). Early information selection proceeds by biasing units representing relevant stimulus features, whereas late selection is achieved by biasing units representing relevant semantic dimensions. Often these models also include components of dual-route models (e.g., the direct link in Phaf et al., 1990). However, they usually lack the idea of a race between early and late selection. Moreover, due to a gradually increasing bias, selectivity in these models changes with time. Thus, these models are rather similar to Cohen et al.'s (1992) model for the flanker task.

Models that bear more resemblance to our model have been proposed by Hunt and Lansman (1986) and by Roelofs (2003). These researchers combined a symbolic production system with features of the connectionist framework. Stimuli are transformed into semantic representations and into a response by production rules. Because the rules are consistently applied in accord with the current goal, only relevant stimulus features drive response selection via this route. However, representations can also be activated by spreading activation across associated nodes. Because associations are rather strong between words and reading responses, words interfere with color naming.

The models of Hunt and Lansman (1986) and Roelofs (2003) share some crucial properties with our model. First, response selection proceeds by a race between a slow but reliable goal-directed process (the productions) and a fast but less reliable automatic process (spreading activation). Second, this race is interactive in a similar way as in our model. The automatic process wins either the race and produces a response (which is often the wrong one for incongruent stimuli), or it affects the speed and accuracy of the controlled process. Third, each process has its own stimulus selection mechanism. Whereas the selection of goal-relevant stimulus information for the controlled process is based on semantic features, which can be considered late selection, the stimulus selection for the automatic process is mainly based on perceptual filtering, at least in Roelofs's (2003) model, which corresponds to early selection. Thus, these models realize very similar ideas as the DSTP model, but they do so within a different framework. How far these models could also account for the present data remains unclear, as they have not yet been fit to distributional data.

Limits and Perspectives

The DSTP model of selective attention introduced in the present article extends existing (formal) models, especially those developed in the area of spatial attention, by a rather general mechanism of late stimulus selection. The presence of both an early and a late stimulus selection mechanism, and the specific way they compete, accounts for relatively complex and detailed attentional phenomena, as the fit of the model to distributional data collected with the flanker task demonstrates. However, the model is also limited. First of all, because it is implemented in the framework of diffusion processes, it applies primarily to speeded responses in forced-choice situations. Moreover, it is rather general and abstract. As a consequence, the model has to leave several details unspecified. For instance, in the present article we assumed that early selection proceeds by spatial filtering, which can be represented by attentional weights. Although we can use our model to estimate how these weights vary across different condition, it says nothing about how the attentional weights are determined. In this sense the DSTP model is not a model of the low-level details of spatial attention. To model these details one would have to include additional mechanisms as specified, for instance, by TVA (Bundesen, 1990) or the CODE theory of visual attention (CTVA; Logan, 1996).

A similar abstractness holds for our late selection process. Formally, it is represented by a diffusion process that comes to a decision at some point in time, which, in turn, changes the rate of the response selection process. In the present article we interpreted this process in the sense that it selects a single item from the display in a categorical manner, and that thereafter the selected item exclusively drives the response selection process. These assumptions, however, are somewhat vague and need to be validated in future work.

However, the fact that some mechanisms of our model are abstract also opens a wide range of potential applications. A potential application of our model is to generate a more detailed picture of those mechanisms that persons use to adapt their selectivity on a trial-by-trial basis to changing conditions. That such a local adaptation can take place has been shown by corresponding sequential effects (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Gratton et al., 1992) and by effects of adaptation after stimulus onset (e.g., Lehle & Hübner, 2008). Current conflict-control theories (e.g., Botvinick et al., 2001; Botvinick, Cohen, & Carter, 2004) assume that early selectivity is adjusted in response to the monitored conflict on the previous trial. By fitting our model to appropriate data one might find that the degree of late selection is adjusted as well. This is only one example of a broad class of conceivable applications of the DSTP model in future research.

References

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44,* 310–329.

Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics, 51,* 145–162.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8,* 539–546.

Brent, R. P. (1973). *Algorithms for function minimization without derivatives.* Englewood Cliffs, NJ: Prentice-Hall.

Broadbent, D. E. (1958). *Perception and communication.* London, England: Pergamon Press.

Brown, P., & Fera, P. (1994). Turning selective attention failure into selective attention success. *Canadian Journal of Experimental Psychology, 48,* 25–57.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response: Linear ballistic accumulation. *Cognitive Psychology, 57,* 153–178.

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

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences, 43,* 345–370.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and nonconscious information processing* (pp. 452–476). Cambridge, MA: MIT Press.

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology, 105,* 239–269.

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 529–553.

Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model

of perceptual grouping by proximity. *Perception & Psychophysics, 53,* 403–421.

Crump, M., Gong, Z., & Milliken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic Bulletin & Review, 13,* 316–321.

de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001, March 2). The role of working memory in visual selective attention. *Science, 291,* 1803–1806.

De Jong, R., Liang, C. C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus–response correspondence. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 731–750.

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

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology, 41,* 260–274.

Diederich, A., & Colonius, H. (1991). A further test of the superposition model for the redundant-signals effect in bimodal detection. *Perception & Psychophysics, 50,* 83–86.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 448–456.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87,* 272–300.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143–149.

Eriksen, B. A., Eriksen, C. W., & Hoffman, J. E. (1986). Recognition memory and attentional selection: Serial scanning is not enough. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 476–483.

Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics, 14,* 155–160.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics, 25,* 249–263.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225–240.

Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology, 58,* 1–48.

Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1). New York, NY: Wiley.

Fernandez-Duque, D., & Knight, M. (2008). Cognitive control: Dynamic, sustained, and voluntary influences. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 340–355.

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1003–1012.

Gegenfurtner, K. R. (1992). PRAXIS: Brent's algorithm for function minimization. *Behavior Research Methods, Instruments, & Computers, 24,* 560–564.

Goolkasian, P., & Bojko, A. (2001). Location constancy and its effect on visual selection. *Spatial Vision, 14,* 175–199.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121,* 480–506.

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin,

E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 331–344.

Harms, L., & Bundesen, C. (1983). Color segregation and selective attention in a nonsearch task. *Perception & Psychophysics, 33,* 11–19.

Heath, R. A. (1992). A general nonstationary diffusion process model for two choice decision making. *Mathematical Social Sciences, 23,* 283–309.

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General, 136,* 217–240.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. *Perception & Psychophysics, 23,* 1–11.

Hommel, B. (1998). Automatic stimulus–response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1368–1384.

Hübner, R., & Backer, G. (1999). Perceiving spatially inseparable objects: Evidence for feature-based object selection not mediated by location. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1556–1567.

Hunt, E., & Lansman, M. (1986). Unified model of attention and problem solving. *Psychological Review, 93,* 446–461.

Johnston, W. A., & Dark, V. J. (1982). In defense of intraperceptual theories of attention. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 407–421.

Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General, 107,* 420–435.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman (Ed.), *Varieties of attention* (pp. 29–61). London, England: Academic Press.

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology, 43,* 711–742.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility: A model and taxonomy. *Psychological Review, 97,* 253–270.

Kramer, A. F., & Hahn, S. (1995). Splitting the beam: Distribution of attention over noncontiguous regions of the visual field. *Psychological Science, 6,* 381–386.

Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception & Psychophysics, 50,* 267–284.

LaBerge, D., & Brown, V. (1986). Variations in size of the visual field in which targets are presented: An attentional range effect. *Perception & Psychophysics, 40,* 188–200.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96,* 101–124.

LaBerge, D., Brown, V., Carter, M., & Bash, D. (1991). Reducing the effects of adjacent distractors by narrowing attention. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 65–76.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review, 107,* 227–260.

Lamberts, K., & Kent, C. (2008). The time course of object-feature retrieval in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 688–695.

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.

Lehle, C., & Hübner, R. (2008). On-the-fly adaptation of selectivity in the flanker task. *Psychonomic Bulletin & Review, 15,* 814–818.

Leuthold, H., Sommer, W., & Ulrich, R. (1996). Partial advance information and response preparation: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General, 3,* 307–323.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 219–234.

Liu, Y. S., Holmes, P., & Cohen, J. D. (2008). A neural network model of

the Eriksen task: Reduction, analysis, and data fitting. *Neural Computation, 20,* 345–373.

Liu, Y. S., Yu, A., & Holmes, P. (2009). Dynamical analysis of Bayesian inference models for the Eriksen task. *Neural Computation, 21,* 1520–1553.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12,* 523–553.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603–649.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109,* 376–400.

Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology, 55,* 207–234.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108,* 393–434.

Logan, G. D., Withey, M. J., & Cowan, W. B. (1977). Cue search and comparison processes in visual search for letters. *Canadian Journal of Experimental Psychology, 31,* 113–121.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in an Stroop-like task. *Memory & Cognition, 3,* 166–174.

Logan, G. D., Zbrodoff, N. J., & Williamson, J. (1984). Strategies in the color–word Stroop task. *Bulletin of the Psychonomic Society, 22,* 135–138.

Luce, R. D. (1986). *Response times.* New York, NY: Oxford University Press.

MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

Mattler, U. (2006). Distance and ratio effects in the flanker task are due to different mechanisms. *Quarterly Journal of Experimental Psychology, 59,* 1745–1763.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaption effects in the absence of executive control. *Nature Neuroscience, 6,* 450–452.

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron, 42,* 677–686.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110,* 422–471.

Miller, J. (1991). The flanker compatibility effect as a function of visual angle, attentional focus, visual transients, and perceptual load: A search for boundary conditions. *Perception & Psychophysics, 49,* 270–288.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11,* 56–60.

Müller, M. M., & Hübner, R. (2002). Can the spotlight of attention be shaped like a donut? *Psychological Science, 13,* 119–124.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003, July 17). Sustained division of the attentional spotlight. *Nature, 424,* 309–312.

Müller, N. G., Mollenhauer, M., Rösler, A., & Kleinschmidt, A. (2005). The attentional field has a Mexican hat distribution. *Vision Research, 45,* 1129–1137.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco, CA: Freeman.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York, NY: Plenum Press.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354.

Paquet, L., & Lortie, C. (1990). Evidence for early selection: Precuing

target location reduces interference from same-category distractors. *Perception & Psychophysics, 48,* 382–388.

Pashler, H., & Badgio, P. C. (1985). Visual attention and stimulus identification. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 105–121.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology, 52,* 629–651.

Phaf, R. H., Van der Heijden, A. H., & Hudson, P. T. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22,* 273–341.

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

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

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

Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology, 21,* 178–184.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88,* 552–572.

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology, 53,* 195–237.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9,* 347–356.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111,* 333–367.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics, 65,* 523–535.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93,* 180–206.

Ridderinkhof, K. R. (1997). A dual-route processing architecture for stimulus–response correspondence effects. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus–response compatibility* (pp. 119–131). Amsterdam, the Netherlands: Elsevier.

Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), *Attention & Performance XIX: Common mechanisms in perception and action* (pp. 494–519). Oxford, England: Oxford University Press.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110,* 88–125.

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 514–523.

Schmidt, J. R., Crump, M. J., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Conscious Cognition, 16,* 421–435.

Schneider, D. W., & Logan, G. D. (2005). Modeling task switching without switching tasks: A short-term priming account of explicitly cued performance. *Journal of Experimental Psychology: General, 134,* 343–367.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and control processing and attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 1–27). Orlando, FL: Academic Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84,* 1–66.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Schwarz, W. (1989). A new model to explain the redundant-signals effect. *Perception & Psychophysics, 46,* 498–500.

Schwarz, W. (1994). Diffusion, superposition and the redundant targets effects. *Journal of Mathematical Psychology, 38,* 504–520.

Servan-Schreiber, D., Bruno, R. M., Carter, C. S., & Cohen, J. D. (1998). Dopamine and the mechanisms of cognition: Part I. A neural network model predicting dopamine effects on selective attention. *Biological Psychiatry, 43,* 713–722.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Shih, S. I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review, 109,* 260–305.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44,* 408–463.

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102,* 503–532.

Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 506–526.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18,* 415–429.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, 40*(A), 201–237.

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

Tyler, C. W., & Kontsevich, L. L. (1995). Mechanisms of stereoscopic processing: Stereoattention and surface perception in depth reconstruction. *Perception, 24,* 127–153.

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed–accuracy tradeoff. *Journal of Mathematical Psychology, 46,* 704–715.

Wendt, M., Kluwe, R. H., & Vietze, I. (2008). Location-specific versus hemisphere-specific adaptation of processing selectivity. *Psychonomic Bulletin & Review, 15,* 135–140.

Wertheimer, M. (1922). Untersuchungen zur Lehre von der Gestalt. I: Prinzipielle Bemerkungen [Studies on the doctrine of the Gestalt: I. Principal comments]. *Psychologische Forschungen, 1,* 47–58.

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

Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 135–149.

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 700–717.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology, 44,* 41–61.

Appendix A

Formulas of the Dual-Stage Two-Phase Model

This appendix provides the formulas of the dual-stage two-phase (DSTP) model. In most cases we derive only the equations for correct responses (Response A) under the condition that the target (Stimulus C) is selected. Because the equations for the other cases are analogue, they can simply be obtained by substitution.

Basic Definitions

A: Boundary for Response A

$-B$: Boundary for Response B

C: Boundary for Target C

$-D$: Boundary for Flanker D

μ: Rate or drift of a given diffusion process and condition

σ^2: Diffusion coefficient or variance of a given diffusion process and condition

*RS*1: Response selection process in Phase 1

*RS*2: Response selection process in Phase 2

SS: Stimulus selection process

We start with a standard Wiener diffusion process (for a primer of diffusion processes see Smith, 2000). The boundary for a correct Response A is denoted by A and that for an error Response B by $-B$. Response A or B is selected if the state of the diffusion process hits boundary A before $-B$, or boundary $-B$ before A, respectively. The time until the process hits one of the boundaries for the first time is called the *first passage time* (FPT). The formulas for the probabilities $p(\text{A})$ and $p(\text{B})$ and for the corresponding expectation of the FPT, $E(T \mid \text{A})$, given Response A, and $E(T \mid \text{B})$, given Response B, can be found elsewhere (e.g., Ashby, 2000; Ratcliff, 1978; Ratcliff & Smith, 2004). These sources also provide formulas for the conditional density functions $f(t \mid \text{A})$ and $f(t \mid \text{B})$ of the FPT and for the corresponding cumulative distribution functions $F(t \mid \text{A})$ and $F(t \mid \text{B})$.

(Appendices continue)

DSTP Diffusion Model

Response Probabilities

In the formal version of the DSTP model we distinguish three diffusion processes and two phases of processing. In the first phase there are two diffusion processes running in parallel and racing against each other (cf. Usher et al., 2002). One process, $RS1$, represents the first part of the response selection process, and the other process, SS, represents the stimulus-selection process. Phase 1 terminates when $RS1$ hits boundary A or $-B$, or when a stimulus is selected. We denote the corresponding random variables representing the FPTs by T_{RS1} and T_{SS}, respectively. Furthermore, the event that $RS1$ wins, that is, that $T_{RS1} < T_{SS}$, is denoted by W_{RS1}, and the event that SS wins is denoted by W_{SS}. If Response A or B is selected by $RS1$, we call this Event A_1 or B_1, respectively. If $RS1$ wins, the corresponding response is triggered and the trial is finished. If SS wins, the response selection process can change its rate and then continues in Phase 2. The response selection process in this phase is represented by an individual diffusion process which is called $RS2$. If Response A or B is selected by $RS2$, we call this Event A_2 or B_2, respectively. If a response symbol is written without index, it represents the event that the response was selected by $RS1$ or by $RS2$. It should be mentioned that, where response selection proceeds across the two phases, the corresponding overall diffusion process is stationary in Phase 1 and in Phase 2 but can be nonstationary across phases. This characteristic is similar to that of other models (cf. Busemeyer & Diederich, 2002; Heath, 1992; W. Schwarz, 1994).

First, we consider the case that a response is selected in Phase 1.

Case W_{RS1}. We start by calculating the probability $p(W_{RS1})$ that $RS1$ wins, that is, the probability that a response is selected before a stimulus is selected. To calculate this probability we need the probability that SS has not yet arrived at a boundary (C or $-D$) by time $t > 0$. This probability corresponds to the survivor function $1 - F_{SS}(t)$, where $F_{SS}(t)$ is the cumulative distribution function of T_{SS}. We further need the density function $f_{RS1}(t)$ of the FPT for $RS1$, which is simply the density of the FPT for a diffusion process, where we use the parameters of the response selection process in Phase 1. With these functions we can compute the probability that $RS1$ wins the race by

$$p(W_{RS1}) = p(T_{RS1} < T_{SS}) = \int_0^\infty f_{RS1}(t)[1 - F_{SS}(t)]dt. \quad (1)$$

The probability that $RS1$ wins, under the condition that Response A_1 is selected, is

$$p(W_{RS1} \mid A_1) = p(T_{RS1} < T_{SS} \mid A_1) =$$

$$\int_0^\infty f_{RS1}(t \mid A_1)[1 - F_{SS}(t)]dt. \quad (2)$$

With these formulas we can compute the probabilities of Response A_1 and Response B_1, respectively, under the condition that $RS1$ wins the race against SS. The corresponding equations are (see Bayes' theorem)

$$p(A_1 \mid W_{RS1}) = \frac{p(W_{RS1} \mid A_1)p(A_1)}{p(W_{RS1} \mid A_1)p(A_1) + p(W_{RS1} \mid B_1)p(B_1)}, \quad (3)$$

$$p(B_1 \mid W_{RS1}) = 1 - p(A_1 \mid W_{RS1}). \quad (4)$$

Case W_{SS}. If $RS1$ and SS race against each other, then it is also possible for SS to win, that is, for a stimulus to be selected before a response is selected. This event W_{SS} occurs with probability $1 - p(W_{RS1})$. The probabilities that SS selects the Stimulus C and the Stimulus D, respectively, conditioned on the event that SS wins the race, can be calculated analogously to those for Responses A and B (see Equations 3 and 4), that is,

$$p(C \mid W_{SS}) = \frac{p(W_{SS} \mid C)p(C)}{p(W_{SS} \mid C)p(C) + p(W_{SS} \mid D)p(D)}, \quad (5)$$

$$p(D \mid W_{SS}) = 1 - p(C \mid W_{SS}). \quad (6)$$

At time T_{ss} when SS wins, $RS1$ is in a certain nonterminated state $X_{RS1}(T_{ss})$ between $-B$ and A. Thus, response selection is unfinished, and we consider this situation as the starting point of Phase 2. In Phase 2, response selection continues and is represented by process $RS2$, whose starting value is the nonterminated state $X_{RS1}(T_{ss})$ of $RS1$, and whose rate might differ from that of $RS1$. Here, we assume that the rate of $RS2$ depends on whether Stimulus C or Stimulus D was selected.

To obtain the density function of the starting value of $RS2$, we have to derive the density of the nonterminated state X_{RS1} of $RS1$ at time T_{ss}. The distribution of this state can be computed by means of a function provided by Ratcliff (1980, 2006):

$$\rho(x, t) = \exp\left(\frac{\mu x}{\sigma^2}\right) \cdot \sum_{n=1}^\infty \frac{2}{(A + B)} \sin\left(\frac{n\pi B}{(A + B)}\right) \sin\left(\frac{n\pi(x + B)}{(A + B)}\right)$$

$$\exp\left[-\frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \frac{n^2\pi^2\sigma^2}{(A + B)^2}\right)t\right]. \quad (7)$$

On the basis of this function we can define the function $\rho_t(t) = \int_{-B}^A \rho(x, t)dx$, which gives the probability that the process is still within its boundaries at time t (i.e., is still a nonterminated process). In our case, this function is identical to the survivor function of $RS1$, that is, $\rho_t(t) = 1 - F_{RS1}(t)$. We can use this function to derive the density function h of state x of $RS1$ at a given time t, which is

$$h(x \mid t) = \frac{\rho(x, t)}{\rho_t(t)} = \frac{\rho(x, t)}{1 - F_{RS1}(t)}, \quad -B < x < A. \quad (8)$$

For a given time t, which determines the density of the start value, the probability that Response A_2 is selected, given Stimulus C, is

$$p(A_2 \mid C, t) = \int_{-B}^A p(A_2 \mid C, x)h(x \mid t)dx. \quad (9)$$

In our model the starting time t in Equation 9 represents the time at which SS wins, that is, it is a random variable with a specific distribution. Thus, the density of the starting value is two-dimensional. To compute this density we also need the density function of the FPT for process SS, given that SS wins the race and a specific stimulus is selected. For instance, when C is selected, we have to derive the density function $f_{SS}(t \mid C, W_{SS})$. This function is given by

$$f_{SS}(t \mid C, W_{SS}) = \frac{1}{p(T_{SS} < T_{RS1} \mid C)} f_{SS}(t \mid C)[1 - F_{RS1}(t)].$$

(10)

The corresponding cumulative distribution function is

$$F_{SS}(t \mid C, W_{SS}) = \int_0^t f_{SS}(y \mid C, W_{SS}) dy.$$

(11)

This formula gives the probability that the race is finished by time t, given that SS wins and Stimulus C is selected.

The density $f_{SS}(t \mid C, W_{SS})$ can now be combined with Equation 8 to obtain the two-dimensional density function of the nonterminated state of $RS1$, conditioned on the events that SS wins and Stimulus C is selected:

$$q(x, t \mid C, W_{SS}) = h(x \mid t) f_{SS}(t \mid C, W_{SS}).$$

(12)

The response selection process $RS2$ begins with a starting value defined by this two-dimensional density or by the corresponding density for Stimulus D. By combining the different equations, we can now calculate the probabilities of Response A_2, given a certain selected stimulus. This probability is obtained by conditioning on the starting value, whose distribution depends on the finishing time of the race between $RS1$ and SS, and on the selected stimulus. For instance, the probability of Response A_2, given Stimulus C, is

$$p(A_2 \mid C, W_{SS}) = \int_0^\infty \int_{-B}^A p(A_2 \mid C, x) h(x \mid t) f_{SS}(t \mid C, W_{SS}) dx dt.$$

(13)

The corresponding probability for Response B_2 is $p(B_2 \mid C, W_{SS}) = 1 - p(A_2 \mid C, W_{SS})$, and the probabilities of A_2 and B_2, conditioned on Stimulus D, are calculated analogously.

We can now calculate the probability $p(A)$ of Response A by combining the probabilities of the three considered cases: the case that $RS1$ wins and selects Response A_1; the case that SS wins, Stimulus C is selected, and $RS2$ selects Response A_2; and the case that SS wins, Stimulus D is selected, and $RS2$ selects Response A_2. By combining these three possible events we obtain

$$p(A) = p(A_1 \mid W_{RS1}) p(W_{RS1}) + [p(A_2 \mid C, W_{SS}) p(C \mid W_{SS})$$
$$+ p(A_2 \mid D, W_{SS}) p(D \mid W_{SS})] p(W_{SS}).$$

(14)

To compute the probability of A under different conditions, it is important to remember that the result of the stimulus selection has different effects. First, whether C or D is selected could influence the distribution of the starting value of $RS2$. Furthermore, the selected stimulus determines the rate of $RS2$. If we apply the model to the flanker task, for instance, then, in case of a congruent stimulus (i.e., if C and D are mapped onto the same response), it seems reasonable to assume that the rate does not depend on the selected stimulus. For incongruent stimuli (i.e., if C and D are mapped onto opposite responses), however, target (C) selection should lead to a higher rate than flanker (D) selection. Moreover, it might even be assumed that the selection of the flanker leads to a negative drift for $RS2$.

Density Functions

For deriving the formulas for the density functions of the FPTs, we first consider the functions that result from the race between the processes SS and $RS1$ (cf. Usher et al., 2002). Equation 10 already gives the density function of the FPT for process SS, conditioned on the events that SS wins and Stimulus C is selected. The density function of the FPT for $RS1$, conditioned on the events that $RS1$ wins and Response A_1 is selected, is analogous:

$$f_{RS1}(t \mid W_{RS1}, A_1) = \frac{1}{p(W_{RS1} \mid A_1)} f_{RS1}(t \mid A_1)[1 - F_{SS}(t)]. \quad (15)$$

To derive the density function of the FPT for process $RS2$, we first consider the case for a given time t_{ss}, at which SS wins and selects a certain stimulus. If we assume that Stimulus C is selected, the distribution function of the starting value is $h(x \mid A_2, C, t_{ss})$, that is, the distribution of the nonterminated state of $RS1$ at time t_{ss}, conditioned on the events that Stimulus C is selected and Response A_2 is selected. Thus, the density function for $RS2$ for this case is given by

$$f_{RS2}(t \mid A_2, C, t_{ss})$$
$$= \begin{cases} 0, & \text{for } t \le t_{ss}. \\ \int_{-B}^A f_{RS2}(t - t_{ss} \mid A_2, C, x) h(x \mid A_2, C, t_{ss}) dx & \text{for } t > t_{ss} \end{cases}$$

(16)

The density function of the starting values, conditioned on Response A_2, is

$$h(x \mid A_2, C, t_{ss}) = \frac{p(A_2 \mid C, x) h(x \mid t_{ss})}{p(A_2 \mid C, t_{ss})}. \quad (17)$$

If we insert Equation 17 into Equation 16, we obtain

$$f_{RS2}(t \mid A_2, C, t_{ss})$$
$$= \begin{cases} 0, & \text{for } t \le t_{ss}. \\ \dfrac{1}{p(A_2 \mid C, t_{ss})} \displaystyle\int_{-B}^A f_{RS2}(t - t_{ss} \mid A_2, C, x) h(x \mid t_{ss}) \\ \qquad\qquad p(A_2 \mid C, x) dx & \text{for } t > t_{ss} \end{cases}$$

(18)

(Appendices continue)

The next step is to assume that the starting time of $RS2$ is itself a random variable whose density is conditioned on the selected stimulus and the selected response. The starting time of $RS2$ is the finishing time of the race between $RS1$ and SS, under the condition that SS wins with a certain stimulus. Thus, the density $f_{RS2}(t \mid A_2, C)$ of the FPT for $RS2$, conditioned on the events that Stimulus C and Response A_2 are selected, is

$$f_{RS2}(t \mid A_2, C, W_{SS}) = \int_0^t \frac{f_{SS}(y \mid A_2, C, W_{SS})}{p(A_2 \mid C, y)}$$

$$\int_{-B}^A f_{RS2}(t - y \mid A_2, C, x)h(x \mid y)p(A_2 \mid C, x)dxdy. \quad (19)$$

By expressing (see Bayes' theorem) the function $f_{SS}(t \mid A_2, C, W_{SS})$ as

$$f_{SS}(t \mid A_2, C, W_{SS}) = \frac{p(A_2 \mid C, t)f_{SS}(t \mid C, W_{SS})}{p(A_2 \mid C, W_{SS})}, \quad (20)$$

and by substituting it in Equation 19, we finally obtain

$$f_{RS2}(t \mid A_2, C, W_{SS}) = \frac{1}{p(A_2 \mid C, W_{SS})}\int_0^t f_{SS}(y \mid C, W_{SS})$$

$$\int_{-B}^A f_{RS2}(t - y \mid A_2, C, x)h(x \mid y)p(A_2 \mid C, x)dxdy. \quad (21)$$

The same function can be used for the case where D is selected.

Now, to compute the density f_{dp} of the FPT for our DSTP model, given Response A, we have to combine the densities of the three possible events. However, we need the probability that $RS1$ wins the race against SS, given that A is selected. By using Equations 1, 3, and 14, we obtain

$$p(W_{RS1} \mid A) = \frac{p(A_1 \mid W_{RS1})p(W_{RS1})}{p(A)}. \quad (22)$$

Thus, we have

$$f_{dp}(t \mid A) = p(W_{RS1} \mid A)f_{RS1}(t \mid A_1, W_{RS1})$$
$$+ p(W_{SS} \mid A)[f_{RS2}(t \mid A_2, C, W_{SS})p(C \mid A_2, W_{SS})$$
$$+ f_{RS2}(t \mid A_2, D, W_{SS})p(D \mid A_2, W_{SS})] \quad (23)$$

with

$$p(C \mid A_2, W_{SS}) = \frac{p(A_2 \mid C, W_{SS})p(C \mid W_{SS})}{p(A_2 \mid W_{SS})}. \quad (24)$$

The probability $p(D \mid A_2)$ is defined in an analogous way.

Expectancies

For deriving the formulas for computing the expected FPT for our model, we first consider the expected FPT $T_1 = T_{RS1}$, given Response A_1 for the case that $RS1$ wins the race. The expected FPT under this condition is

$$E(T_1 \mid W_{RS1}, A_1) = \frac{1}{p(W_{RS1} \mid A_1)}\int_0^\infty t \cdot f_{RS1}(t \mid A_1)[1 - F_{SS}(t)]dt. \quad (25)$$

When SS wins the race, we have a diffusion process with a variable starting value. Let us first consider the case with a fixed starting time t_{ss} for $RS2$. Under this condition the expected FPT T_2 under A_2, if C is selected, is $t_{ss} + E(T_{RS2} \mid W_{SS}, A_2, C)$, that is,

$$E(T_2 \mid A, C, t_{ss}) = t_{ss} + \frac{1}{p(A_2 \mid C, t_{ss})}$$

$$\int_{-B}^A E(T_{RS2} \mid A_2, C, x)h(x \mid t_{ss})p(A_2 \mid C, x)dx. \quad (26)$$

We now take the starting time t_{ss} as a random variable. Analogous to the density we obtain

$$E(T_2 \mid A_2, C, W_{SS}) = \int_0^\infty f_{SS}(t \mid A_2, C, W_{SS})E(T_2 \mid A_2, C, t)dt. \quad (27)$$

If we insert Equation 26 and reformulate the density of the stimulus selection, then we get

$$E(T_2 \mid A_2, C, W_{SS})$$

$$= \frac{1}{p(A_2 \mid C, W_{SS})}\int_0^\infty f_{SS}(t \mid C, W_{SS})[t \cdot p(A_2 \mid C, t)$$

$$+ \int_{-B}^A E(T_{RS2} \mid A_2, C, x)h(x \mid t)p(A_2 \mid C, x)dx]dt. \quad (28)$$

The expected FPT for our DSTP model, given Response A, can be computed in a similar way to the densities:

$$E(T \mid A) = p(W_{RS1} \mid A)E(T_1 \mid A_1, W_{SS})$$
$$+ p(W_{SS} \mid A)[E(T_2 \mid A_2, C, W_{SS})p(C \mid W_{SS}, A_2)$$
$$+ E(T_2 \mid A_2, D, W_{SS})p(D \mid W_{SS}, A_2)]. \quad (29)$$

Cumulative Distribution Functions

The cumulative distribution functions of the DSTP model can be computed by integrating its densities. However, it might be easier to compute them analogously to the densities. We start by considering the distribution of $RS1$. In this case we have to integrate. It is given by

$$F_{RS1}(t \mid W_{RS1}, A_1) = \frac{1}{p(W_{RS1} \mid A_1)}\int_0^t f_{RS1}(y \mid A_1)[1 - F_{SS}(y)]dy. \quad (30)$$

Next, we consider the cases when SS wins the race. In these cases we have a diffusion process with a variable starting point. Analogous to the densities we have

$$F_{RS2}(t \mid A_2, C, W_{SS}) = \frac{1}{p(A_2 \mid C, W_{SS})} \int_0^t f_{SS}(y \mid C, W_{SS})$$

$$\int_{-B}^{A} F_{RS2}(t - y \mid A_2, C, x) h(x \mid y) p(A_2 \mid C, x) dx dy. \quad (31)$$

Combined we have

$$F_{dp}(t \mid A) = p(W_{RS1} \mid A) F_{RS1}(t \mid A_1, W_{RS1})$$
$$+ p(W_{SS} \mid A)[F_{RS2}(t \mid A_2, C, W_{SS}) p(C \mid W_{SS}, A_2)$$
$$+ F_{RS2}(t \mid A_2, D, W_{SS}) p(D \mid W_{SS}, A_2)]. \quad (32)$$

Appendix B

Method and Results of Experiments 1, 2, and 3

Experiment 1

Method

Participants. Twenty participants (ranging from 20 to 28 years; 7 men, 13 women) with normal or corrected-to-normal vision participated in the study. Participants were recruited at the Universität Konstanz and were paid €8 per hour.

Apparatus. The stimuli were presented on a 19-in. monitor with a resolution of 1280×1024 pixels. A personal computer served for controlling stimulus presentation and response registration.

Stimuli. The stimulus set consisted of the numerals from 2 to 9. The height of the numerals subtended a visual angle of $1.27°$ at a viewing distance of 45 cm, and their width was about $0.89°$, depending on the specific digit. Items were presented in white on a black background. The target always appeared at the central position of the screen. Flankers consisted of two copies of a numeral, which were presented left and right of the target at eccentricities of $1.27°$ and $2.8°$ in the narrow and wide conditions, respectively.

Procedure. Each trial started with a fixation cross presented for 400 ms, which was followed by a blank screen for 600 ms (cue-stimulus interval) and by a subsequent stimulus array presented for 165 ms. The task was to judge the parity (odd or even) of the target. The participants had to press a left key for "even" and a right key for "odd" with the index and middle finger of their right hand, respectively. Flankers were congruent on half of the trials (i.e., they had the same parity as the target) and were incongruent on the other half (i.e., had the opposite parity as the target). One second after the response, the next trial began. Errors were signaled by short tones.

Each block consisted of 64 trials. After a preliminary practice session, the participants worked through 28 test blocks in a 2-hr session. Pairs of blocks with the wide stimulus spacing alternated with pairs of blocks with the narrow stimulus spacing throughout the experiment. The order was balanced across participants. In all there were 448 trials for each condition.

Results

Response times (RTs). The latencies of correct responses were analyzed by a two-factor analysis of variance (ANOVA) for repeated measures on the factors stimulus spacing (wide vs. narrow) and congruency (congruent vs. incongruent). The analysis revealed sig-

nificant main effects of stimulus spacing, $F(1, 19) = 69.2, p < .001$, and of congruency, $F(1, 19) = 72.5, p < .001$. However, there was also a reliable interaction between these two factors, $F(1, 19) = 51.8, p < .001$. It indicates that the congruency effect was larger in the narrow condition than in the wide condition (29 ms vs. 7 ms).

Error rates. The mean error rate was 10.9%. The error rates were subjected to an ANOVA of the same type as for the RTs. It revealed significant main effects of stimulus spacing, $F(1, 19) = 27.6, p < .001$, and of congruency, $F(1, 19) = 91.1, p < .001$. However, there was also a reliable interaction between these two factors, $F(1, 19) = 52.5, p < .001$. It indicates that the congruency effect was larger in the narrow condition than in the wide condition (9.35% vs. 2.33%).

Experiment 2

Method

Twenty participants (ranging from 20 to 36 years; 7 men, 13 women) with normal or corrected-to-normal vision participated in the study. Participants were recruited at the Universität Konstanz and were paid €8 per hour. The same apparatus and the same stimuli (narrow condition) as in our first experiment were used. Here, however, there were three blocked stimulus-position conditions: The stimuli could occur either at the center of the screen (1-position-central), at an eccentricity of $1.27°$ (center of the screen to center of the target) to the left or to the right (2-position-lateral), or equally often at all three positions (3-position).

Seven blocks for the 1-position-central condition, seven blocks for the 2-positions-lateral condition, and 14 blocks for the 3-positions condition with 64 trials each were mixed and administered in a 2-hr session. In the 2-positions-lateral and 3-positions conditions the stimuli were presented with equal frequency at each of the possible locations. In all, we had 448 trials for the 1-position-central and the 2-positions-lateral conditions and 896 trials for the 3-position condition. Congruent and incongruent stimuli occurred with equal frequency.

Results

RTs. The data were first analyzed with a three-factor ANOVA for repeated measures. The factors were eccentricity (central vs.

(Appendices continue)

lateral), uncertainty (low vs. high), and congruency (congruent vs. incongruent). The analysis revealed a significant main effect of eccentricity, $F(1, 19) = 168$, $p < .001$. Responses were faster to central stimuli than to lateral stimuli (474 ms vs. 511 ms). However, there was also a reliable interaction between eccentricity and uncertainty, $F(1, 19) = 29.9$, $p < .001$. It indicates that the difference between the eccentricity conditions was larger (45 ms vs. 28 ms) if uncertainty was low (i.e., in the 1-position-central and 2-positions-lateral conditions) than when it was high (i.e., in the 3-positions-central and 3-positions-lateral conditions). There was also a reliable main effect of congruency, $F(1, 19) = 262$, $p < .001$. However, congruency interacted significantly with uncertainty, $F(1, 19) = 5.48$, $p < .05$, and with eccentricity, $F(1, 19) = 13.5$, $p < .01$. The first interaction indicates that the congruency effect was larger under high than under low uncertainty (42 ms vs. 35 ms), whereas the latter interaction indicates that the congruency effect was larger for the lateral positions than for the central position (44 ms vs. 32 ms).

Error rates. The mean error rate was 12.6%. The error rates were subjected to an ANOVA of the same type as for the RTs. It revealed significant main effects of eccentricity, $F(1, 19) = 6.06$, $p < .05$, and of congruency, $F(1, 19) = 128$, $p < .001$. However, there was also a significant interaction between these two factors, $F(1, 19) = 25.4$, $p < .001$. The congruency effect was larger for the lateral positions than for the central position (11.9% vs. 7.71%).

Experiment 3

Method

Sixteen participants (ranging from 19 to 36 years; 6 men, 10 women) with normal or corrected-to-normal vision participated in the study. Participants were recruited at the Universität Konstanz and were paid €8 per hour. Apparatus, stimuli, and procedure were similar to the 2-positions condition in our second experiment. Here, however, the cue-stimulus interval was reduced to 200 ms.

Moreover, we had two block types: In one type (80%-congruent) 80% of the flankers were congruent, whereas in the other type (20%-congruent) 20% of the flankers were congruent. On the remaining trials the flankers were incongruent. After one training block for each type, the participants alternated through nine blocks of 80 trials for each type in a single 1.5-hr session, starting with the 80%-congruent type. This resulted in 576 and 144 trials for the majority and minority trial types, respectively. Eight participants, who produced too few errors for a distributional analysis, especially for the congruent stimuli in the 20%-congruent blocks, were invited to a second session some days later. Thus, for these participants we had twice as many data for the analysis.

Results

RTs. The latencies of correct responses were analyzed by an overall two-factor ANOVA for repeated measures on the factors block type (80%-congruent vs. 20%-congruent) and congruency (congruent vs. incongruent).

The analysis revealed significant main effects of congruency, $F(1, 15) = 255$, $p < .001$, and of block type, $F(1, 15) = 22.2$, $p < .001$. However, there was also a significant interaction between these two factors, $F(1, 15) = 6.53$, $p < .05$. The congruency effect was larger in the 80%-congruent condition than in the 20%-congruent condition (42 ms vs. 34 ms).

Error rates. The mean error rate was 11.8%. The error rates were subjected to an ANOVA of the same type as that for the RTs. The analysis revealed significant main effects of congruency, $F(1, 15) = 170$, $p < .001$, and of block type, $F(1, 15) = 40.3$, $p < .001$. However, there was also a significant interaction between the two factors, $F(1, 15) = 48.5$, $p < .001$. The congruency effect was larger in the 80%-congruent condition than in the 20%-congruent condition (15.4% vs. 8.16%).