



HAL
open science

Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws

Mathieu Servant, Anna Montagnini, Boris Burle

► To cite this version:

Mathieu Servant, Anna Montagnini, Boris Burle. Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognitive Psychology*, 2014, 72, pp.162-195. 10.1016/j.cogpsych.2014.03.002 . hal-01308931

HAL Id: hal-01308931

<https://hal.science/hal-01308931>

Submitted on 28 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



ELSEVIER

Contents lists available at ScienceDirect

Cognitive Psychology

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



Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws



Mathieu Servant^{a,b}, Anna Montagnini^b, Borís Burle^{a,*}

^a *Laboratoire de Neurosciences Cognitives, Fédération de Recherche 3C, case C, Aix-Marseille Université, CNRS, 3 Place Victor Hugo, 13331 Marseille cedex 3, France*

^b *Institut de Neurosciences de la Timone, Aix-Marseille Université, CNRS, Faculté de Médecine, 27 boulevard Jean Moulin, 13005 Marseille, France*

ARTICLE INFO

Article history:

Accepted 28 March 2014

Available online 21 April 2014

Keywords:

Diffusion model

Decision-making

Conflict tasks

Piéron's law

Wagenmakers–Brown's law

Cognitive control

ABSTRACT

Formal models of decision-making have traditionally focused on simple, two-choice perceptual decisions. To date, one of the most influential account of this process is Ratcliff's drift diffusion model (DDM). However, the extension of the model to more complex decisions is not straightforward. In particular, conflicting situations, such as the Eriksen, Stroop, or Simon tasks, require control mechanisms that shield the cognitive system against distracting information. We adopted a novel strategy to constrain response time (RT) models by concurrently investigating two well-known empirical laws in conflict tasks, both at experimental and modeling levels. The two laws, predicted by the DDM, describe the relationship between mean RT and (i) target intensity (Piéron's law), (ii) standard deviation of RT (Wagenmakers–Brown's law). Pioneering work has shown that Piéron's law holds in the Stroop task, and has highlighted an additive relationship between target intensity and compatibility. We found similar results in both Eriksen and Simon tasks. Compatibility also violated Wagenmakers–Brown's law in a very similar and particular fashion in the two tasks, suggesting a common model framework. To investigate the nature of this commonality, predictions of two recent extensions of the DDM that incorporate selective attention mechanisms were simulated and compared to the experimental results. Both models predict Piéron's law and the violation of Wagenmakers–Brown's law by compatibility. Fits of the models to the RT distributions and accuracy data

* Corresponding author.

E-mail address: boris.burle@univ-amu.fr (B. Burle).

allowed us to further reveal their relative strengths and deficiencies. Combining experimental and computational results, this study sets the groundwork for a unified model of decision-making in conflicting environments.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Over the past 40 years, the fundamental process of making decisions on the basis of sensory information, known as perceptual decision-making, has grown up to an extensive field of research. The interest has increased in part due to the introduction of the sequential sampling framework (for reviews, see Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Ratcliff & Smith, 2004). To make a decision, it is assumed that the brain accumulates samples of sensory evidence until an absorbing choice boundary is reached. The inherent noise in both the physical stimulus and the neural signal makes the process stochastic, potentially leading to an incorrect choice. The rate of approach to a boundary is called *drift rate*, and depends on the quality of the extracted sensory evidence. The boundary is hypothesized to be under subjective control, and can be modulated depending on timing demands. A higher boundary criterion will require greater evidence accumulation, leading to slower and more accurate decisions. The interaction between drift rate and choice criteria has an obvious property: it provides an integrated account of both response time (RT) and accuracy in choice laboratory experiments.

The drift diffusion model (DDM) developed by Ratcliff and coworkers (Ratcliff, 1978; Ratcliff & Rouder, 1998) belongs to this theoretical frame. The model was originally developed to explain simple two-choice decisions in terms of psychologically plausible processing mechanisms, and has proven to account for a large range of paradigms (for a review, see Ratcliff & McKoon, 2008). However, its extension to more complex decisions is not straightforward and is currently the object of an intense field of research in both experimental psychology (e.g., Hübner, Steinhauser, & Lehle, 2010; Leite & Ratcliff, 2010; Smith & Ratcliff, 2009; Stafford, Ingram, & Gurney, 2011; White, Brown, & Ratcliff, 2011; White, Ratcliff, & Starns, 2011) and neuroscience (e.g., Churchland, Kiani, & Shadlen, 2008; Resulaj, Kiani, Wolpert, & Shadlen, 2009). The present study aims to evaluate whether the DDM can be extended to conflicting situations, and contributes to this emerging field.

1.1. The drift diffusion model: basic architecture and mathematical properties

As other sequential sampling models, the DDM posits that RT is the sum of two components, a non-decision time and a decision-related time. The decision process takes the form of an accumulation of evidence delimited by two boundaries representing alternative choices. The starting point of the diffusion depends on prior expectations, and can be located everywhere on the axis joining the two alternatives, being closer to the more expected alternative. In each moment, the incremental evidence is the difference between sensory inputs supporting choice 1 versus 2. This difference is a random variable which follows a Gaussian distribution, with mean μ (drift rate) and variance σ^2 (diffusion coefficient). The combination of sensory evidence into a single variable and its linear stochastic accumulation over time present an interesting property. If the diffusion is discretized, then the process becomes a random walk and is formally equivalent to the sequential probability ratio test (SPRT; Wald, 1947). SPRT is optimal in the sense that it minimizes expected decision time for any given accuracy level, and maximizes accuracy for a given decision time (Wald & Wolfowitz, 1948). Bogacz et al. (2006) have argued that optimality may be a hallmark of human *cognitive control*, the ability to adapt information processing from moment to moment depending on current goals. According to this view, the DDM may provide a privileged framework to study such control processes, and offers an interesting departure point to approach decision-making in conflicting situations.

Two properties are predicted by the DDM when task difficulty (drift rate) is manipulated. Those properties have so consistently been observed in both detection¹ and choice experiments that psychologists have proposed them to be psychological laws. First, the mean and standard deviation (SD) of RT distributions increase at approximately the same rate when drift rate declines. Empirically, the linear relationship between the mean and SD of RT distributions holds for a broad range of paradigms and generally leads to very high correlations for each individual (Pearson's $r > .85$; Luce, 1986; Wagenmakers & Brown, 2007; hereafter referred to as Wagenmakers–Brown's law). Second, the chronometric function predicted by the DDM when the two alternatives are equiprobable is an hyperbolic tangent function of the following form:

$$\text{Mean RT} = \frac{a}{\mu} \tanh\left(\frac{a\mu}{\sigma^2}\right) + \text{Ter}$$

where a , μ , and σ^2 are respectively the boundary, drift rate, and diffusion coefficient of the diffusion process (Ratcliff, 1978). *Ter* is the non-decision time. For a suprathreshold range of stimulus intensities, this function mimics Piéron's law (see Palmer, Huk, & Shadlen, 2005, Experiment 3). Piéron's law states that mean RT decreases as a power function of the intensity of a stimulus according to:

$$\text{Mean RT} = \alpha I^{-\beta} + \gamma$$

where α is a scaling parameter, I represents stimulus intensity, γ the asymptotic RT, and β determines the rate of decay of the curve (Piéron, 1913). Although initially investigated in the context of detection tasks (e.g., Chocholle, 1940), Piéron's law has proven to hold in choice experiments (Palmer et al., 2005; Pins & Bonnet, 1996; Stafford et al., 2011; van Maanen et al., 2012). In conclusion, Piéron and Wagenmakers–Brown's laws are consistent with the diffusion framework, and may reflect a general tendency of human decision-makers to approach optimal behavior.

1.2. Conflict paradigms

Besides “simple” situations, one often has to make decisions in a multiple stimuli environment, only some of those stimuli being relevant for the task at hand. One source of paradigms designed to study such situations are so-called conflict tasks. Empirical findings in these tasks converge toward an apparent stimulus–response (S–R) compatibility effect. Responses are slowed down and less accurate when a task-irrelevant aspect of a multi-attribute stimulus is associated to the response opposite to that of the relevant aspect (Hommel, 2011; Kornblum, Hasbroucq, & Osman, 1990). The interfering association can be physical, conceptual, or artificially created by task instructions. Examples of such conflict tasks are the Stroop (Stroop, 1935), the Eriksen flanker (Eriksen & Eriksen, 1974), and the Simon (Simon & Small, 1969). The Stroop task requires participants to report the ink color of a word string. The word denotes a color that can be either identical to the ink (e.g., the word “blue” printed in blue ink) or different (e.g., the word “blue” printed in red ink). In the Eriksen task, subjects give a manual response to a central symbolic target (e.g., a right response for the letter S and a left response for the letter H) flanked by distracters calling for the same (SSS) or opposite (HSH) response. Finally, in the classical version of the Simon task, subjects are requested to press a right or left button in response to the color of a lateralized stimulus. Conflict arises when stimulus position and response side do not correspond.

The existence of interference effects demonstrates that performance is suboptimal. Because the standard DDM implements an optimal decision-making strategy (Bogacz et al., 2006), one can hypothesize that it will have difficulties to account for conflicting situations. The present work investigates how conflict tasks interact with Piéron and Wagenmakers–Brown laws, and how recent extensions of the DDM cope with such interactions. Through these investigations, we aim to highlight potential processing similarities and lay the foundation for a unified framework of decision-making in conflicting environments. Two DDM extensions that incorporate selective attention mechanisms are

¹ At first glance, the DDM appears incompatible with detection tasks. However, detection can be conceptualized as a choice between the presence of a particular stimulation versus noise (see van Maanen, Grasman, Forstmann, & Wagenmakers, 2012).

simulated and their predictions with regard to Piéron and Wagenmakers–Brown laws tested against experimental data from two different conflict tasks. A final evaluation of the models is performed by fitting them to the full data sets, taking into account RT distributions and accuracy. While DDM extensions capture critical properties of the two psychological laws, common to both conflict paradigms, they fail to qualitatively reproduce the complete pattern of data. Their relative strengths and deficiencies are further elucidated through their fits.

1.3. Modeling decision-making in conflicting situations with diffusion processes

Distributional analyses in conflict tasks have revealed faster errors than correct responses when S–R are incompatible. Notably, plots of accuracy rates as a function of RT quantile (i.e., *conditional accuracy functions*, CAFs) show a characteristic drop of accuracy for faster RT quantiles in this condition. By contrast, CAFs for compatible trials are relatively flat (Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988; Hübner & Töbel, 2012; White, Brown, et al., 2011; Wylie, Ridderinkhof, Bashore, & van den Wildenberg, 2010; Wylie et al., 2012). Previous studies have indicated that a standard DDM can produce faster errors than correct responses if and only if inter-trial variability in the starting point of the accumulation process is added (Laming, 1968; Ratcliff & Rouder, 1998). However, this additional assumption would lead to equivalent increases in fast errors in the compatible condition, contrary to what is observed. Two recent diffusion developments account for this asymmetry by assuming an increase in attentional selectivity for the relevant stimulus attribute over the course of a trial, whatever the S–R mapping. The improvement of the quality of evidence induces a *time-varying drift rate*. The two models, depicted in Fig. 1, differ regarding whether selective attention operates in a discrete (dual-stage two-phase model of selective attention, DSTP; Hübner et al., 2010) or gradual manner (shrinking-spotlight model, SSP; White, Ratcliff, et al., 2011). In the DSTP, response selection is performed by a diffusion variable with two functionally different phases. The drift rate of the first phase is governed by sensory information passing through an early attentional filter (early selection stage). It is defined as the sum of two component rates, one for the relevant stimulus attribute μ_{rel} and the other for the irrelevant attribute μ_{irrel} (μ_{irrel} is negative in incompatible trials). Because the

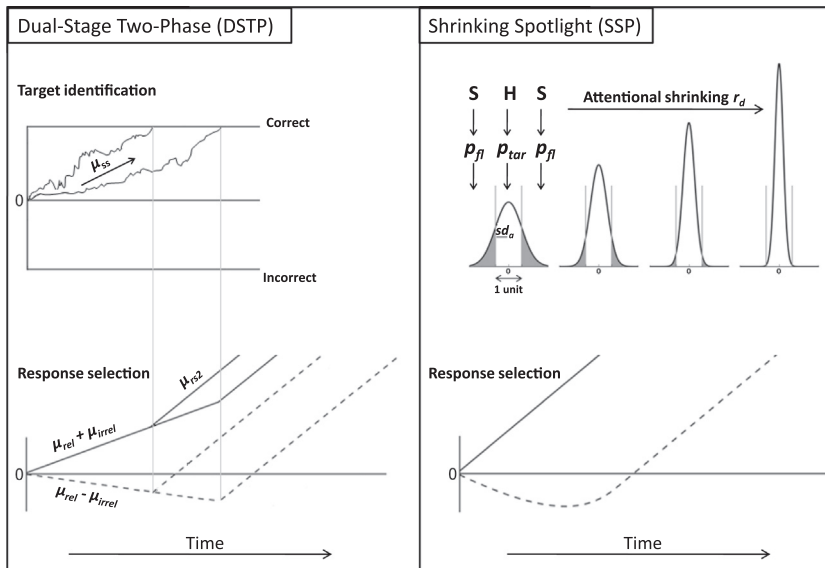


Fig. 1. Basic architectures of the DSTP (left panel) and SSP (right panel) diffusion models. The rate of evidence accumulation for response selection improves discretely in the DSTP and continuously in the SSP (compatible trials: plain lines; incompatible trials: dashed lines). See text for details.

early attentional filter is imprecise, μ_{irrel} often prevails over μ_{rel} , and the net drift rate moves toward the incorrect response boundary in incompatible trials, provoking fast errors. In parallel, a second diffusion variable with drift rate μ_{ss} fulfills the role of target identification (late selection stage). Because two diffusion processes are racing, different scenarios can occur. (i) The response selection variable reaches a boundary before the target identification variable. In this case, the model reduces to a standard DDM, and responses are mainly determined by the irrelevant stimulus attribute. Conversely, a target can be identified before the selection of a response. (ii) If the identification is correct, the drift rate of response selection increases discretely from $\mu_{rel} \pm \mu_{irrel}$ to μ_{rs2} . This second phase of response selection, driven exclusively by the selected stimulus, counteracts early incorrect activations in incompatible trials and explains the improved accuracy of slower responses (see Fig. 1, left panel, for an illustration of this scenario). (iii) If the identification is incorrect, μ_{rs2} is negative, and the model generates a slow perceptual error. Taking the Eriksen task as a working example, Hübner and colleagues showed that their model could account for RT distributions and accuracy under a wide range of experimental conditions. However, the DSTP has been challenged by a more parsimonious single-stage model with a continuous time-varying drift rate. White, Ratcliff, et al. (2011) used the attentional zoom-lens analogy (Eriksen & St James, 1986) as a basic mechanism for weighting sensory evidence over time. Their SSP model was specifically developed to account for spatial attention dynamics in the Eriksen task, and was consequently formalized in a less abstract way compared to the general selective attention framework of the DSTP. The SSP assumes that spatial attention is dispersed early in a trial, allowing influence from the flankers, and progressively narrows toward the central target, whatever the S–R compatibility (see Fig. 1, right panel). The attention weight attributed to the target (flankers) is modeled as the integral of a unitary Gaussian distribution with standard deviation sd_a , over a region of space corresponding to the target (flankers). Importantly, sd_a decreases at a linear rate r_d . In every time step, the perceptual input of the target p_{tar} and each flanker p_{fl} is weighted by the allocated quantity of attention, and the resulting evidence defines the evolving drift rate. p_{fl} is positive in compatible trials and negative in incompatible trials. For a standard Eriksen task, the model assumes that each item provides the same quantity of evidence p ($p = p_{tar} = p_{fl}$). Under this assumption, the drift rate in compatible trials is constant (the attention weights always sum to 1). The situation is different in incompatible trials where the drift rate is initially directed toward the incorrect boundary, triggering fast errors, and progressively turns toward the correct boundary as attention shrinks. White and colleagues demonstrated that this simple model provides a better fit performance compared to the DSTP in the Eriksen task, although strong mimicry has been noticed. Hübner and Töbel (2012) recently showed that the superiority of the SSP is actually tied to specific experimental situations. Indeed, the fits of both models are virtually indiscernible for the RT distributions of correct responses. The discrepancy concerns the dynamic of errors in the incompatible condition. The SSP predicts an improvement of accuracy that is too fast, a problem attenuated when the proportion of fast errors is low. However, the divergence is small and further emphasizes model mimicry. Further computational details regarding the spotlight component of the SSP are provided in Appendix A.

An important property of the DSTP and SSP models is that they predict larger RT mean and SD for the incompatible compared to the compatible S–R condition, that is, a consistent RT moment ordering. The shrinking mechanism of the SSP is assumed to operate similarly across S–R mappings, and the drift rate for incompatible stimuli gradually converges toward that of compatible stimuli, but never surpasses it.² Because the diffusion coefficient remains constant, this scheme necessarily leads to a wider spread of RT for the incompatible condition (see Schwarz & Miller, 2012, for a similar reasoning based on another continuous time-varying drift rate scheme). The same logic applies to the DSTP, with a discrete convergence of drift rates toward μ_{rs2} . Although the onset and sign of μ_{rs2} are conditional on the late selection stage, this additional flexibility does not challenge, on average, the consistent RT moment ordering between compatibility conditions. While this statistical pattern is characteristic of the Stroop (Pratte, Rouder, Morey, & Feng, 2010) and Eriksen tasks (Hübner et al., 2010; Ridderinkhof, Scheres, Oosterlaan, & Sergeant, 2005; White, Ratcliff, et al., 2011), the reversed ordering has consistently been observed in the standard version of the Simon task (Burle, Possamai, Vidal, Bonnet, & Hasbroucq,

² More precisely, the drift rate converges toward a limit equal to the perceptual input of the target.

2002; Pratte et al., 2010; Ridderinkhof, 2002; Schwarz & Miller, 2012). That is, the incompatible condition is associated with the largest mean and the *smallest* SD, which violates Wagenmakers–Brown’s law. This singularity led researchers to propose that the Simon effect may be incompatible with the diffusion framework (Pratte et al., 2010; Schwarz & Miller, 2012). Given the success of time-dependent diffusion processes in modeling the Eriksen task, such an assumption would mean that decision-making draws upon qualitatively different mechanisms depending on the nature of the conflicting situation.

1.4. Piéron and Wagenmakers–Brown laws as a window on decision-making mechanisms

As introduced above, Piéron and Wagenmakers–Brown laws are hallmarks of a standard DDM with constant drift rate. In their studies, neither Hübner et al. nor White et al. (Hübner & Töbel, 2012; Hübner et al., 2010; White, Brown, et al., 2011; White, Ratcliff, et al., 2011) explored properties of their model when the perceptual intensity of the relevant stimulus attribute is manipulated. Simulations of the SSP and DSTP, presented in Section 2, aimed to determine whether Piéron and Wagenmakers–Brown laws still hold under the assumption of time-varying decision evidence. To our knowledge, the two laws have never been concurrently investigated in conflict tasks. An exception is found in a recent study by Stafford et al. (2011). Those researchers manipulated the intensity of colors in a standard Stroop task. Five suprathreshold color saturation levels were presented in an intermixed fashion. In each compatibility condition, mean RT and color discriminability scaled according to Piéron’s law. Interestingly, the two factors combined in an additive fashion. Results remained similar when the word and the color were spatially separated (i.e., separate Stroop task). Section 3 extends those findings by providing an empirical test of Piéron and Wagenmakers–Brown laws in Eriksen and Simon tasks. The Eriksen task was naturally chosen insofar as the DSTP and SSP models have specifically been tested on it. The Simon task was also introduced because we could anticipate a violation of Wagenmakers–Brown’s law. To allow a direct comparison between the two experiments, we used the standard Simon task and a version of the Eriksen task in which subjects have to discriminate the color of a central circle while ignoring the color of flanking circles (Davranche, Hall, & McMorris, 2009). The perceptual intensity of the target could thus be varied along the same color saturation dimension. Color saturation was manipulated within a highly controlled perceptual color space while keeping constant any other aspect of the display. Notably, the color saturation of the flanking circles in the Eriksen task remained invariable (maximum saturation level), just as the eccentricity of the targets in the Simon

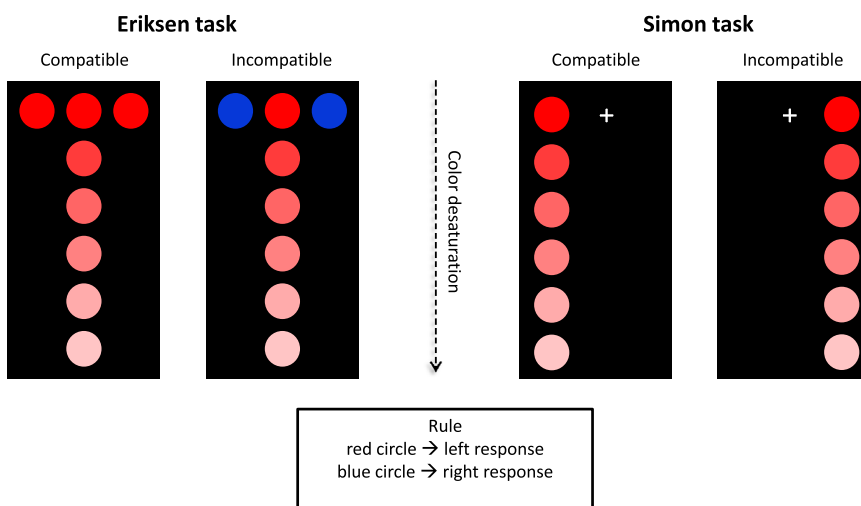


Fig. 2. Stimuli used in the present experiments (only the variation of color saturation for the red stimulus is shown as example).

task (see Fig. 2). Six color saturation levels were chosen to span a wide range of color intensities, and were presented in a randomized fashion. The design offers sufficient experimental conditions to concurrently investigate Piéron and Wagenmakers–Brown laws.

2. Model simulations

Because the SSP is intractable mathematically (Ratcliff, 1980), both models were simulated using a random walk numerical approximation (Diederich & Busemeyer, 2003; Ratcliff & Tuerlinckx, 2002) and a 1 ms time step. Our simulations aimed at quantifying the mean and SD of decision times (DT) for each compatibility condition when the perceptual intensity of the relevant stimulus attribute is manipulated while that of the irrelevant attribute remains constant. To obtain reliable estimates of SD, 100,000 trials per condition were simulated. As a parametric baseline, we used the best-fitting group parameters for each model reported by White, Ratcliff, et al. (2011) from Experiment 1 (standard Eriksen task) and assumed unbiased starting points of diffusion processes.

2.1. SSP predictions

The SSP model has five free parameters: a (boundary separation), Ter (non-decision time), p (perceptual input of any item in the display), sd_a (standard deviation of the Gaussian distribution), and r_d (attentional shrinking rate). The parametric baseline was $a = 0.129$, $p = 0.383$, $sd_a = 1.861$, $r_d = 0.018$ (see White, Ratcliff, et al., 2011, Table 2). Ter was set to zero. To manipulate independently the perceptual intensity of the target and the flankers, the perceptual input parameter p was decomposed into the input for the target p_{tar} and the input for each flanker p_{fl} . p_{tar} decreased from 0.383 to 0.183 in steps of 0.01, corresponding to different levels of perceptual intensity. p_{fl} was equal to 0.383 (maximum perceptual intensity). All the remaining model parameters were kept constant. Fig. 3A represents the simulated SSP prediction for the mean and SD of DT across conditions. Wagenmakers–Brown's law holds for the perceptual factor, but is strongly violated by S–R compatibility. Focusing on mean DT also reveals an increase of the compatibility effect as p_{tar} decreases, because it takes more time for the decision process to overcome incorrect activations. The Piéron's like behavior of the predicted chronometric functions is obvious from Fig. 3B, where the relationship between p_{tar} and mean DT is plotted in a log–log space. The approximate linearity is diagnostic of a power function analogous to Piéron's law.

2.2. DSTP predictions

The DSTP model has seven free parameters: a (boundary separation for the response selection process), Ter (non-decision time), c and μ_{ss} (boundary separation and drift rate for the target identification process), μ_{rel} (component rate for the relevant stimulus attribute), μ_{irrel} (component rate for the irrelevant stimulus attribute), and μ_{rs2} (drift rate for the second phase of response selection). The parametric baseline was $a = 0.128$, $c = 0.177$, $\mu_{ss} = 1.045$, $\mu_{rel} = 0.108$, $\mu_{irrel} = 0.241$, and $\mu_{rs2} = 0.414$ (see White, Ratcliff, et al., 2011, Table 2). Ter was set to zero. The early selection stage of the DSTP is not modeled. Perceptual inputs receive early attention weights giving rise to the component rates for the relevant and irrelevant stimulus attributes, μ_{rel} and μ_{irrel} . We thus decomposed μ_{rel} assuming that it is the product of $p_{tar} = 0.383$ (perceptual input of the target) and an attention weight of 0.282. This gives $\mu_{rel} = 0.108$, which is the best-fitting value reported by White and colleagues. p_{tar} was manipulated in the same way as the SSP, decreasing from 0.383 to 0.183 in steps of 0.01. μ_{irrel} remained constant (0.241). Because the perceptual manipulation necessarily affects the identification of the target, μ_{ss} also decreased from 1.045 (best-fitting value) to 0.445 in steps of 0.03. Fig. 3C and D show the resulting predictions. Similar to the SSP, the DSTP predicts Piéron and Wagenmakers–Brown laws for each compatibility mapping separately. The compatibility effect also increases when the perceptual intensity of the target decreases, because both early and late selection mechanisms are reduced.

Under difficult target selection conditions (e.g., narrow spacing between target and flankers), Hübner et al. (2010) observed that μ_{rs2} increases to keep performance at a reasonable level, at least when target selection difficulty is manipulated blockwise. Whether this compensatory mechanism

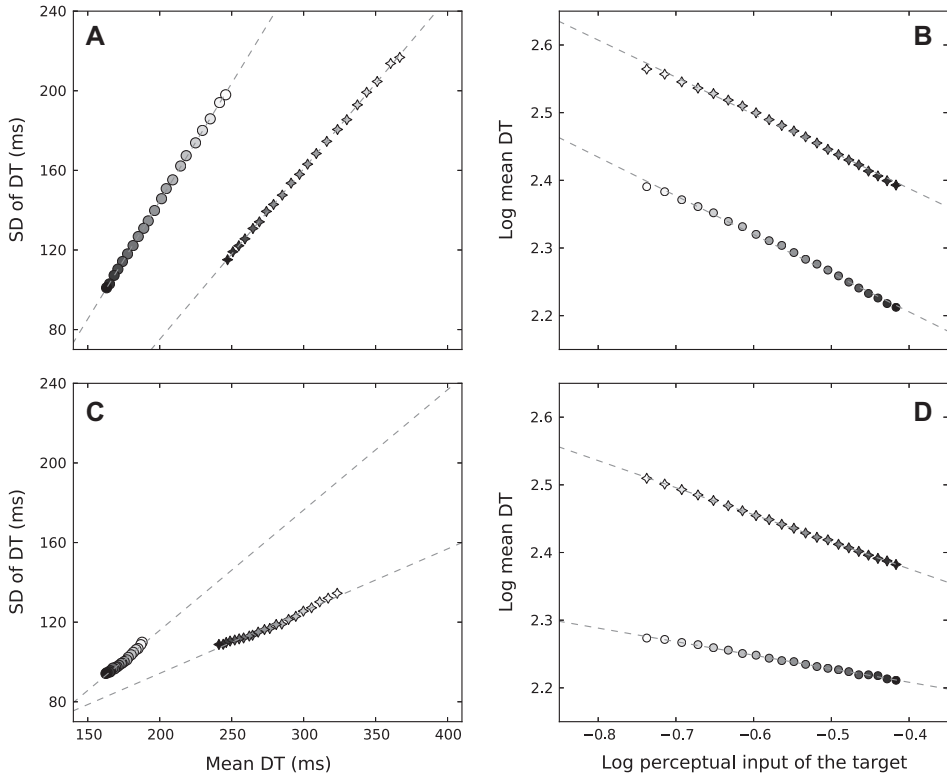


Fig. 3. Predictions from the SSP (A, B) and DSTP (C, D) diffusion models for the Wagenmakers–Brown (left panels) and the log-transformed Piéron (right panels) laws when the perceptual intensity of the target decreases while that of the irrelevant stimulus attribute remains maximal. Circles represent compatible conditions, stars represent incompatible conditions. Perceptual input levels of the target are symbolized by gray shading. Also shown are lines of best fit for each compatibility mapping (dashed lines). Left panels: SD of DT versus mean DT (both in milliseconds). Right panels: chronometric functions displayed in a log–log space.

holds for randomized designs is uncertain. In [Appendix B](#), we provide an additional simulation of the DSTP, identical to the previous one, except that μ_{rs2} increases from 0.414 to 0.490 as target intensity decreases. This slight parametric variation produces a curvilinear shape for the relationship between the mean and SD of DT within each compatibility condition (see [Fig. B.1](#)). Since, anticipating our empirical findings, we have a strong linear relationship for target intensity, a constant μ_{rs2} provides a more parsimonious model and a better description of this aspect of the data.

2.3. Discussion

The present simulations uncover similar chronometric properties of the SSP and DSTP models. Piéron and Wagenmakers–Brown laws are predicted for each compatibility condition separately along with a super-additive interaction between target intensity and compatibility. These predictions are largely similar to those of a standard DDM ([Stafford et al., 2011](#)). A major difference should be emphasized, however: the linear relationship between the mean and SD of RT distributions, proposed to be a psychological law, is broken by the compatibility factor. In line with our theoretical analysis of time-varying drift rate dynamics (see introduction, [Section 1.3](#)), the SSP and DSTP models also produce a consistent DT moment ordering between compatibility conditions, and this is true for every target intensity level (as can be observed, in [Fig. 3A](#) and [C](#), by comparing point and star markers with the

same gray shading). Those singularities represent a benchmark against which empirical data will be contrasted.

3. Experiments

3.1. Experiment 1: Eriksen task

3.1.1. Method

3.1.1.1. *Participants.* Twelve students (*Mean age* = 23.7 years, *SD* = 4.4, 7 female) from the University of Aix-Marseille completed experiment 1 and were paid 10 €/h. They were naive with respect to the purpose of the experiment and reported to have normal or corrected-to-normal vision and normal color vision. This experiment was approved by the ethical committee of the Aix-Marseille University, and by the “Comité de Protection des Personnes Sud Méditerranée 1” (approval n° 1041). Participants gave their informed written consent according to the declaration of Helsinki.

3.1.1.2. *Stimuli and apparatus.* Subjects were tested individually in a dark room (~ 0.08 cd/m²). They were seated in a comfortable chair 150 cm in front of a CRT color monitor with a refresh rate of 100 Hz. At this distance, 1 cm on the screen corresponded to approximately 0.38° of visual angle. Stimulus presentation and collection of data were controlled by Psychopy (Peirce, 2007). Special attention was paid to the manner in which Psychopy utilizes the vertical refresh rate/sync of the monitor to ensure RT data was not influenced by the vertical blank interval. Stimuli were red and blue circles (radius = 0.24°) presented on the horizontal midline of a 12.18° × 9.15° black field. On every trial, a target circle appeared in the center of the field and was flanked by two circles at an eccentricity of 0.8° center to center. We manipulated the color saturation of target circles while keeping their luminance constant. To obtain identical levels of perceptual saturation between red and blue, we used the CIE Lightness Chroma Hue device-independent³ colorimetric space (Commission Internationale de l’Eclairage, 1976), which is a variant of the CIE L*a*b* space specifically designed to accurately map color perception. Chroma quantifies the degree of perceptual saturation across colors. Lightness is a non-linear transformation of luminance. Although the two concepts are different, it is always true that colors with the same lightness will have the same luminance. Six suprathreshold chroma levels (15%, 25%, 35%, 45%, 60%, and 80%) were chosen to span a large range of color intensities. Red (Hue = 30°) and blue (Hue = 280°) colors always had the same lightness (*L* = 51), corresponding to a luminance of approximately 19 cd/m². The chroma level of the flankers was set to 80%, and was never modulated. Colors were calibrated by means of a Brontes colorimeter (Admesy B.V., The Netherlands). Responses were made by the subject pressing either a right or a left button with the corresponding thumb. Button closures were transmitted through the parallel port of the computer to reach high temporal precision. Buttons were arranged on the top of two plastic cylinders (3 cm in diameter, 7 cm in height) serving as handgrips, and the distance between the cylinders was 20 cm.

3.1.1.3. *Procedure.* Subjects performed 24 blocks of 96 trials in a single-session experiment lasting approximately 100 min. Within a block, trials were defined by factorial combination of flanker hue (red or blue), target hue (red or blue), and target chroma (6 saturation levels). Half of the trials were congruent (i.e., same color for target and distracters), and the other half were incongruent (different color). They were pseudo-randomized by Mix software (van Casteren & Davis, 2006) so that first order compatibility sequences (i.e., compatible–incompatible CI, CC, IC, and II) occurred the same number of times. Chroma levels were not paired equally often with each of the possible compatibility sequences, since this would have added too many constraints and could have led to predictability. However, a posteriori analysis showed that chroma levels were fairly well balanced within the compatibility transition matrix. On every trial, an array of three circles was presented, and remained on-screen until the subject responded, with a maximum duration of 1500 ms. The next trial started 1500 ms after

³ Although the CIE LCH is considered a device-independent color space, it needs a white point reference. We chose the standard and widely used CIE D65 illuminant.

stimulus offset. Subjects were instructed to respond as fast and as accurately as possible to the color of the central circle, and to ignore the color of the flanking circles. Half of the subjects gave a left-hand response to a blue target and a right-hand response to a red target. This mapping was reversed for the other half of the subjects. At the beginning of the experiment, subjects performed a practice block similar to the experimental blocks. Practice trials were excluded from analyses.

3.1.1.4. Power function fitting procedure. Luce (1986) proposed that Piéron's law encompasses a non-decision component (the asymptotic RT γ , in the sense that it is the shortest RT that can be achieved) and a decision-related one (the power term). In line with this assumption, Bonnet (1992) found that γ was only sensitive to sensory modalities, and argued that it was tied to non-decisional processes. Similarly, electromyographic evidence suggests that motor execution is insensible to flanker and Simon interferences (Burle et al., 2002; Hasbroucq, Possamai, Bonnet, & Vidal, 1999; Rösler & Finger, 1993). In contrast, S–R compatibility factors are traditionally thought to affect decision-making (Kornblum et al., 1990), and this hypothesis represents the core basis of the DSTP and SSP models. We thus constrained our fitting procedure to produce a unique γ estimate for compatible and incompatible conditions. Other parameters (α , β) were free to vary between compatibility conditions. A loss function was computed between the theoretical power curve and empirical data, and this function was minimized with a SIMPLEX routine (Nelder & Mead, 1965). The initial parameter values were drawn from uniform distributions with boundaries defined by previous fits of Piéron's law to choice data (Stafford et al., 2011; van Maanen et al., 2012). Each power function was fitted several times, best-fitting values serving as a starting point for the next run. Stimulus discriminability was operationalized using chroma parameters. Following Stafford et al. (2011), Piéron's law was fitted both on the group and individual data, and goodness-of-fit was assessed by computing Pearson's correlation coefficient between observed and predicted data.

3.1.1.5. Data analyses. Mean RT and proportions of errors were submitted to an ANOVA with flanker compatibility (compatible, incompatible) and chroma (6 saturation levels) as within-subject factors. An arc-sine transformation was applied to normalize proportions before analysis (Winer, 1971). Greenhouse–Geisser corrections were applied in case of violation of the sphericity assumption (Greenhouse & Geisser, 1959). Other specific analyses will be detailed in the text.

3.1.2. Results

3.1.2.1. RTs and accuracy. Anticipations (responses faster than 100 ms, 0.007%) and trials in which participants failed to respond (0.03%) were discarded. There was a reliable flanker effect on RT ($M = 44.5$ ms), $F(1, 11) = 42.4$, $p < .001$, $\eta_p^2 = 0.79$ (see Table 1). The main effect of chroma was also significant, $F(5, 55) = 60.7$, $p < .001$, $\varepsilon = 0.41$, $\eta_p^2 = 0.85$. Lower chroma levels were associated with slower RT (amplitude of the effect, $M = 58.9$ ms). Importantly, the interaction between chroma and compatibility was not significant, $F(5, 55) = 0.6$, $p = 0.6$, $\varepsilon = 0.5$, $\eta_p^2 = 0.05$. Error rates revealed a similar pattern. We found main effects of compatibility, $F(1, 11) = 17.6$, $p < .005$, $\eta_p^2 = 0.62$, and chroma,

Table 1

Mean response times (ms) and error percentages (parentheses) for each flanker compatibility condition and chroma level for Experiments 1–2.

Experiment and Cond	Chroma levels (%)					
	15	25	35	45	60	80
1. Eriksen						
comp	445 (14.2)	421 (8)	410 (5.3)	400 (4.3)	392 (4.1)	387 (4.7)
incomp	488 (32.6)	472 (22.4)	452 (15.4)	447 (13.2)	434 (11.4)	428 (10.7)
2. Simon						
comp	420 (15.9)	408 (10.1)	390 (7)	379 (5.7)	373 (4.8)	370 (6.8)
incomp	447 (22.8)	428 (14)	414 (9.5)	402 (7.9)	390 (7.8)	388 (6.9)

Note. comp: compatible trials; incomp: incompatible trials.

$F(5,55) = 52.7, p < .001, \varepsilon = 0.5, \eta_p^2 = 0.83$. Error rate was higher in the incompatible condition, and increased as chroma decreased. The interaction between chroma and compatibility failed to reach significance, $F(5,55) = 2.03, p = 0.17, \varepsilon = 0.3, \eta_p^2 = 0.16$.

In order to provide some quantitative support to the plausibility of the null hypothesis of additivity, we further computed a Bayesian ANOVA on mean RT (Rouder, Morey, Speckman, & Province, 2012) with the R package Bayesfactor (Morey & Rouder, 2012). More precisely, we evaluated the ratio of the marginal likelihood of the data given model M_0 (implementing additive effects between compatibility and color saturation) and given model M_1 (including interactive effects between compatibility and color saturation), a ratio known as Bayes factor. The Bayes factor measures the relative change in prior to posterior odds brought about by the data:

$$\frac{\overbrace{p(M_0/Data)}^{\text{posterior odds}}}{\overbrace{p(M_1/Data)}^{\text{posterior odds}}} = \frac{\overbrace{p(M_0)}^{\text{prior odds}}}{\overbrace{p(M_1)}^{\text{prior odds}}} \times \frac{\overbrace{p(Data/M_0)}^{\text{Bayes factor}}}{\overbrace{p(Data/M_1)}^{\text{Bayes factor}}} \quad (1)$$

The Bayes factor for M_0 over M_1 was $BF_{0,1} = 15.1 \pm 0.63\%$, revealing that the data are 15 times more likely under the additive model M_0 compared to the interactive model M_1 . According to a standard scale of interpretation (Jeffreys, 1961), a Bayes factor of about 15 represents *strong* evidence for M_0 .

3.1.2.2. Piéron's law. Fig. 4 displays the best-fitting Piéron's curve for each flanker compatibility condition along with observed mean RT. From a qualitative point of view, Piéron's law seems to describe the data well. This impression is reinforced by very high correlation coefficients between observed and predicted data, both at the group and the individual levels (see Tables 2 and 3).

3.1.2.3. Wagenmakers–Brown's law. Following the methodology of Wagenmakers and Brown (2007), we computed the RT mean and standard deviation (SD) of our 12 experimental conditions for each individual. Fig. 5A displays the resulting scatter plots along with Pearson's r coefficients of correlation and lines of best fit. The r values ranged from 0.49 to 0.96 with a mean of 0.78, and the majority of subjects showed an $r < .85$ (9 out of 12 subjects). The parameters of the linear relationship seem to be influenced by the S–R compatibility factor. This impression is reinforced when the mean and SD of each experimental condition are averaged across subjects (see Fig. 5B).

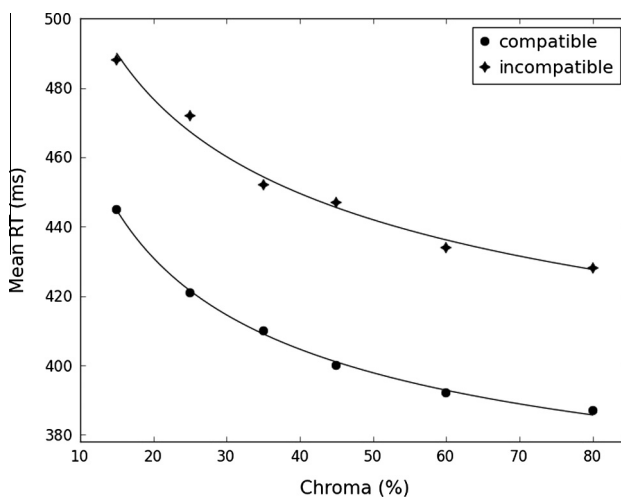


Fig. 4. Piéron's law for each flanker compatibility condition of Experiment 1. Symbols represent empirical mean response time (RT) data, solid lines are best fitting power curves.

Table 2

Fit of Piéron's law to the group data for each compatibility condition, Experiments 1–2.

Experiment and Cond	Correlation coefficient	γ (ms)	α	β
<i>1. Eriksen</i>				
comp	0.999 [*]	337	44	0.47
incomp	0.993 [*]	337	85	0.31
<i>2. Simon</i>				
comp	0.984 [*]	279	83	0.29
incomp	0.993 [*]	279	100	0.28

Note. comp: compatible trials; incomp: incompatible trials.

* Significant $p < .001$.

Table 3

Fit of Piéron's law to the individual data for each compatibility condition, Experiments 1–2.

Experiment and Cond	Correlation coefficients	
	<i>M</i>	<i>SD</i>
<i>1. Eriksen</i>		
comp	0.92	0.12
incomp	0.88	0.11
<i>2. Simon</i>		
comp	0.92	0.11
incomp	0.96	0.04

Note. comp: compatible trials; incomp: incompatible trials.

To try to separate out the effects of random variability from the experimental manipulations, we built a linear mixed effects model (Pinheiro & Bates, 2000). Contrary to general linear model methods, mixed models allow to structure the variance of the observations by modeling random effects. This development leads to more constrained parameter estimates. The models were specified using the R package lme4 (Bates, Maechler, & Bolker, 2012). We estimated p -values by means of Markov chain Monte Carlo (MCMC, Baayen, Davidson, & Bates, 2008). Model selection was performed by computing a Bayesian information criterion (BIC; Schwarz, 1978) that penalizes models according to their complexity (i.e., number of free parameters). The best model is the one with the smallest BIC. Such a model predicted SD of RT based on mean RT and compatibility as fixed factors along with by-subject random intercepts. The interaction term between mean RT and compatibility was removed, because its contribution was not significant and penalized the model. We found main effects of mean RT and compatibility (both MCMC $p < .001$). Controversies exist regarding how model selection should be done and which statistical assessment should be performed (e.g., Barr, Levy, Scheepers, & Tily, 2013; Schielzeth & Forstmeier, 2009). In Appendix C, we provide additional analyses with more complex random effect structures and likelihood ratio tests to assess fixed effects. All analyses converged and confirmed our observations. The compatibility factor violates Wagenmakers–Brown's law by modulating its intercept. The best-fitting parameter for the fixed effect of compatibility indicates that the intercept is lowered by about 10 SD units in the incompatible condition. Note, however, that for each level of chroma, both RT mean and SD are larger in the incompatible than the compatible condition.

3.1.3. Discussion

In agreement with the DSTP and SSP predictions, the results of Experiment 1 show that Piéron and Wagenmakers–Brown laws hold for each compatibility condition separately in an Eriksen task. Linear mixed effects model analyses revealed that the intercept of the linear relationship between RT mean and SD is lowered by the incompatible mapping. However, time-varying diffusion models also predict an effect of compatibility on the slope of the linear law (see Fig. 3). This discrepancy may be linked to the relationship between compatibility and color saturation. We found that the two factors combine additively, as revealed by a Bayesian analysis, while diffusion models predict a super-additive interaction.

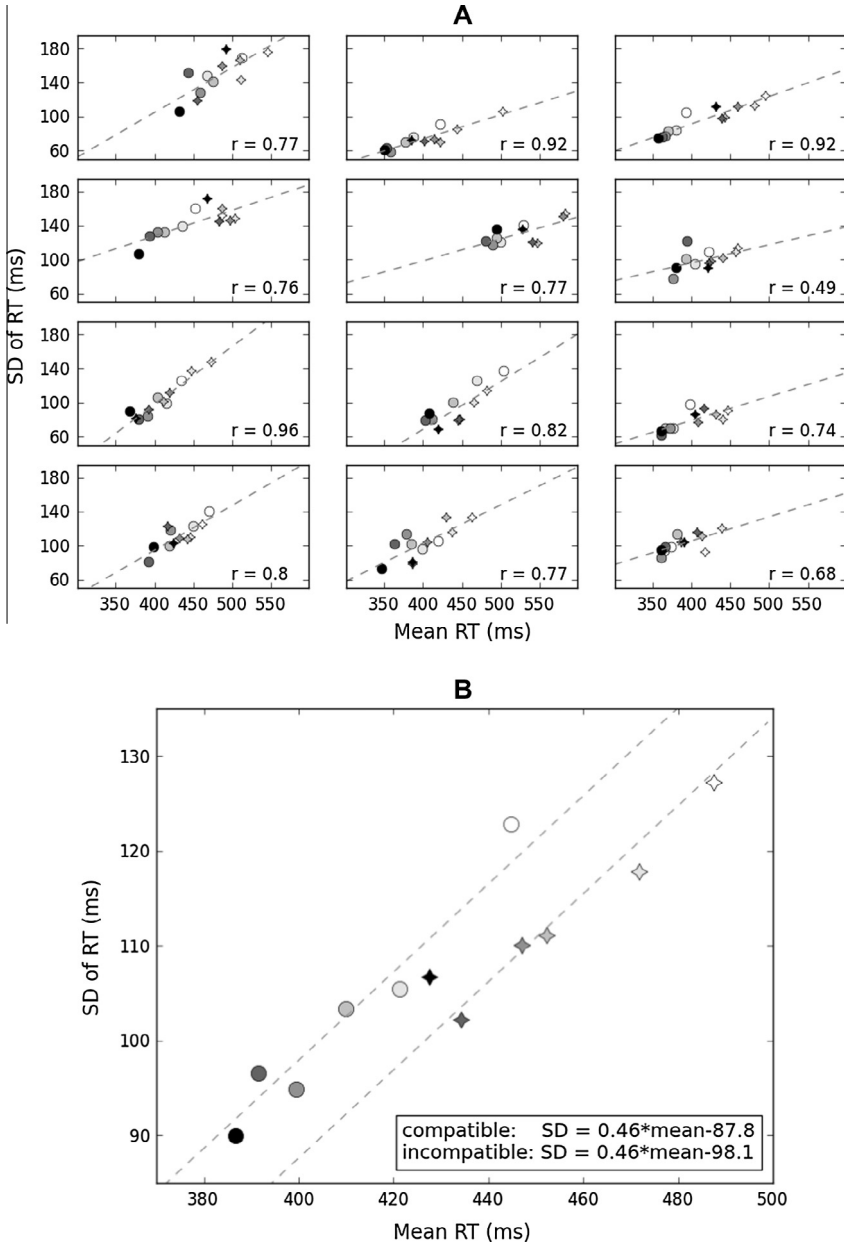


Fig. 5. Standard deviation (SD) of response time (RT) versus mean RT across the 12 experimental conditions from Experiment 1. Circles represent compatible conditions, stars represent incompatible conditions. Chroma levels are symbolized by gray shading. (A) Each subplot displays data from 1 of the 12 participants. Lines of best fit (dashed lines) and Pearson's r coefficient correlations are also shown. (B) Data averaged across subjects. Dashed lines are constructed using the best-fitting parameters of the more parsimonious linear mixed effects model. Their equation is provided in the inset.

The next experiment investigates another conflicting situation, the Simon task, considered to be incompatible with the diffusion framework due to an inconsistent RT moment ordering between

compatibility conditions (Schwarz & Miller, 2012). Consequently, particular attention will be paid to how RT mean and SD scale across experimental conditions.

3.2. Experiment 2: Simon task

3.2.1. Method

3.2.1.1. Participants. Twelve students (Mean age = 23 years, SD = 2.4, 6 female) were recruited from the same pool as Experiment 1 and were paid 10 €/h. None of them was informed in advance about the purpose of the experiment, and none of them participated in the first experiment. All the students reported to have normal or corrected-to-normal vision and normal color vision. This experiment was approved by the ethical committee of the Aix-Marseille University, and by the “Comité de Protection des Personnes Sud Méditerranée 1” (approval n° 1041). Participants gave their informed written consent according to the declaration of Helsinki.

3.2.1.2. Stimuli and apparatus. Stimuli, colors and apparatus were identical to Experiment 1. In each trial, however, only one circle was presented 1.6° to the left or right of the vertical midline. A 0.23° × 0.23° gray cross in the center of the screen served as fixation point. The luminance of the cross was identical to that of the colors (~19 cd/m²).

3.2.1.3. Procedure. Subjects worked through 28 blocks of 96 trials in a single-session experiment lasting approximately 100 min. Within a block, trials were defined by factorial combination of stimulus location (left or right), hue (red or blue) and chroma (6 saturation levels). They were pseudo-randomized in the same way as Experiment 1. A trial started by the presentation of a fixation cross. One second later, a target circle appeared either to the right or to the left of fixation. Stimuli disappeared as soon as a response was emitted, or after a response deadline set to 1000 ms. Subjects were instructed to respond as fast and as accurately as possible to the color of the circle irrespective of its position. Half of the subjects gave a left-hand response to a blue target and a right-hand response to a red target. This mapping was reversed for the other half of the subjects. At the beginning of the experiment, subjects performed a practice block similar to the experimental blocks. Practice trials were excluded from analyses.

3.2.2. Results

3.2.2.1. RTs and accuracy. Anticipations (responses faster than 100 ms, 0.02%) and trials in which participants failed to respond (0.18%) were discarded. There were main effects of compatibility on RT, $F(1, 11) = 70.2$, $p < .001$, $\eta_p^2 = 0.87$ (Simon effect, $M = 21.6$ ms; see Table 1), and chroma, $F(5, 55) = 86$, $p < .001$, $\varepsilon = 0.5$, $\eta_p^2 = 0.89$, (amplitude of the effect, $M = 54.9$ ms). The interaction between chroma and compatibility was not significant, $F(5, 55) = 1.5$, $p = .2$, $\eta_p^2 = 0.1$. Main effects of compatibility and chroma were also reliable on error rates, $F(1, 11) = 13.5$, $p < .005$, $\eta_p^2 = 0.6$ and $F(5, 55) = 52.5$, $p < .001$, $\varepsilon = 0.5$, $\eta_p^2 = 0.83$ respectively. However, there was a slight trend for a compatibility × chroma interaction, $F(5, 55) = 2.4$, $p = .09$, $\varepsilon = 0.5$, $\eta_p^2 = 0.18$. Tukey post hoc tests revealed that the Simon effect was only reliable for 15% ($p < .05$) and 25% ($p < .001$) chroma levels.

A Bayesian ANOVA was further computed on mean RT in the same way as Experiment 1. The data favored the additive model M_0 over the interactive model M_1 by a Bayes factor of $BF_{0,1} = 7.2 \pm 0.61\%$, providing substantial support for additive effects (Jeffreys, 1961).

3.2.2.2. Piéron's law. Best fitting Piéron's law for each compatibility condition and observed mean RT are displayed in Fig. 6. As in Experiment 1, Piéron's law describes the data well. The correlation coefficients between observed and predicted data are very high, both at the group and the individual levels (see Tables 2 and 3).

3.2.2.3. Wagenmakers–Brown's law. The data was analyzed in the same way as Experiment 1. Pearson's r values for each individual are generally lower compared to those observed in the Eriksen task (mean = 0.58, range 0.15–0.95; see Fig. 7A). A rapid look at the averaged data (Fig. 7B) makes clear that Wagenmakers–Brown's law is violated by the compatibility factor. As anticipated, the incompatible

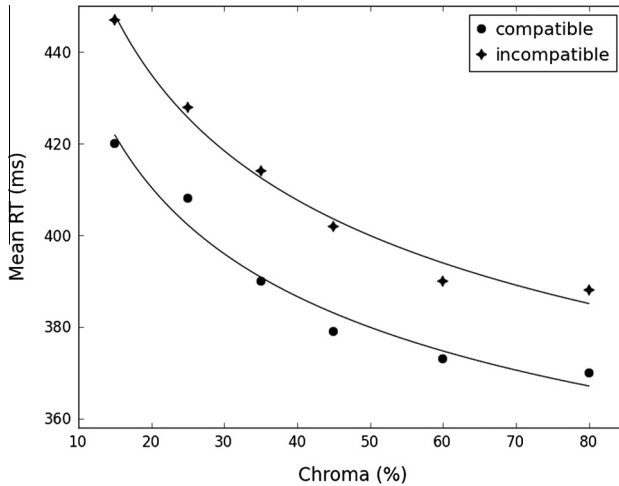


Fig. 6. Piéron's law for each compatibility condition of Experiment 2. Symbols represent empirical mean response time (RT) data, solid lines are best fitting power curves.

condition is associated with a smaller SD of RT compared to the compatible condition for each color saturation level. The linear mixed effects model with the lowest BIC index comprised by-subject random intercepts, and RT mean and compatibility as fixed factors. The interaction term was again removed, because it was not significant and penalized the model. The effects of compatibility and mean RT were reliable (both MCMC $p < .001$). The best-fitting parameter for the fixed effect of compatibility revealed that the intercept of Wagenmakers–Brown's law was lowered by about 15 SD units in the incompatible condition (see [Appendix C](#), for additional analyses leading to similar conclusions).

3.2.3. Discussion

The pattern of results from Experiment 2 is similar to that previously observed in the Eriksen task. Piéron and Wagenmakers–Brown laws hold for each S–R compatibility condition separately. The incompatible mapping lowers the intercept of the linear law by about 15 SD units, but does not affect its slope. Those results provide strong support for a common model framework between Eriksen and Simon tasks, and time-varying diffusion models appear likely candidates. While the DSTP is sufficiently abstract to be extended to different conflict tasks ([Hübner et al., 2010](#)), the SSP was specifically designed for spatial attentional control. However, [White, Ratcliff, et al. \(2011\)](#) hypothesized that the spotlight component of the SSP may also center on a more abstract feature space to account for non-spatial attentional effects in the Eriksen task (e.g., grouping effects). In the Simon task, the irrelevant spatial feature of the stimulus would be gradually suppressed as the spotlight narrows on the relevant color feature.

Two aspects of the data, however, seem to challenge the models. In line with previous studies, we found an inconsistent RT moment ordering between compatibility conditions in the Simon task, but not in the Eriksen (see [Figs. 5B and 7B](#)). Moreover, compatibility and color saturation combined additively in the two conflict tasks. In the next section, we provide a final test of the SSP and the DSTP by fitting them to the RT distributions and accuracy data of the previous experiments. This test is more powerful than the RT mean and SD approach taken so far, and should provide a detailed picture of the relative strengths and deficiencies of the models. We also fit an alternative version of the SSP, proposed post hoc by [White, Ratcliff, et al. \(2011\)](#). This model features a lack of attentional shrinking in the compatible condition, and was motivated by the empirical finding that subjects tend to minimize attentional effort whenever possible. When the perceptual intensity of the target and flankers is similar, as in a standard Eriksen task, each item provides the same quantity of evidence. There is no real advantage of shrinking attention on the target in compatible trials, and a lack of shrinking does

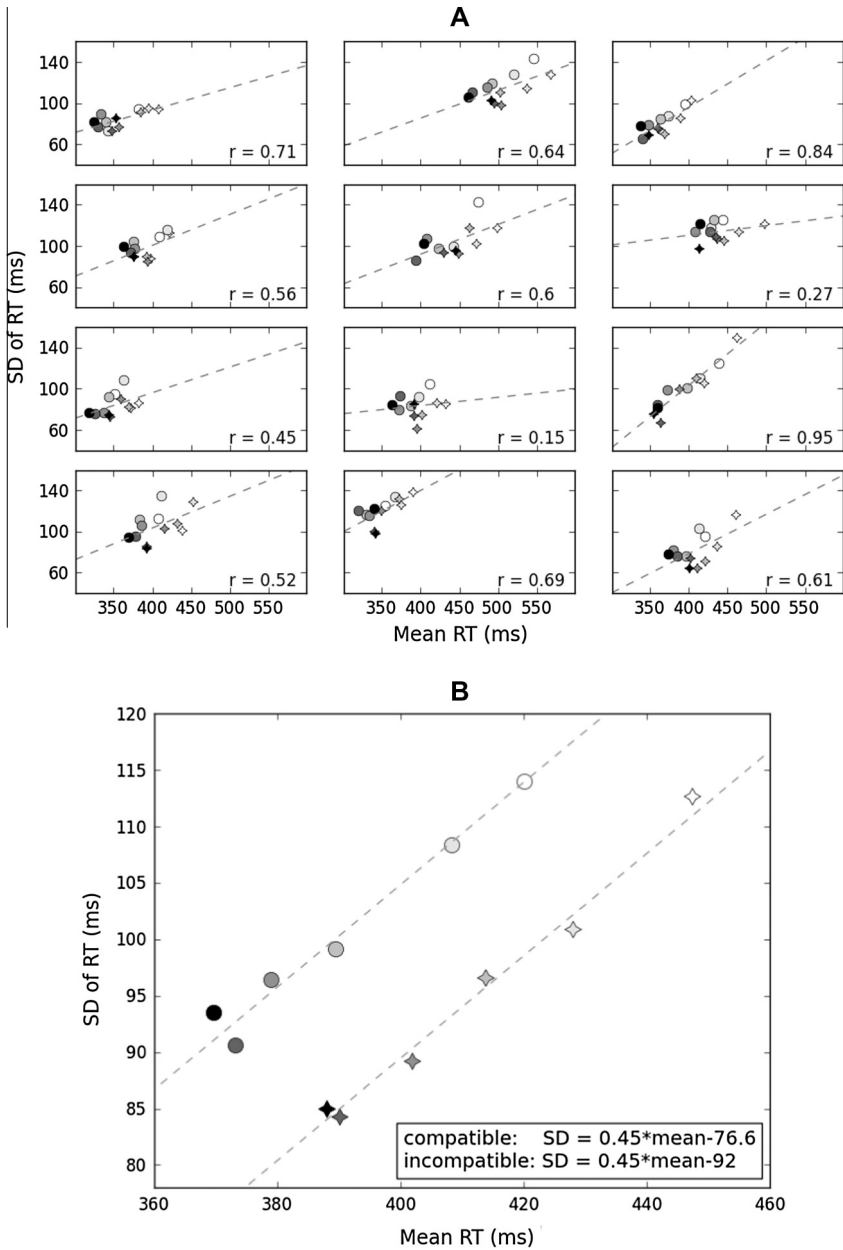


Fig. 7. Standard deviation (SD) of response time (RT) versus mean RT across the 12 experimental conditions from Experiment 2. Circles represent compatible conditions, stars represent incompatible conditions. Chroma levels are symbolized by gray shading. (A) Each subplot displays data from 1 of the 12 participants. Best fit lines (dashed lines) and Pearson's r coefficient correlations are also shown. (B) Data averaged across subjects. Dashed lines are constructed using the best-fitting parameters of the more parsimonious linear mixed effects model. Their equation is provided in the inset.

not alter the model's behavior (the constant drift rate in compatible trials would remain unchanged). This is not true when the perceptual intensity of the target and flankers is manipulated independently. In the original SSP, if $p_{tar} < p_{fl}$, the drift rate in compatible trials would become time-varying and would

progressively converge toward p_{tar} . However, a lack of attentional shrinking would always induce a constant drift rate, partly determined by p_{fl} . There are two interesting properties of this alternative SSP model. First, simulations reveal a pattern that resembles our empirical findings: the incompatible mapping lowers the intercept of Wagenmakers–Brown’s law but does not affect its slope (see [Appendix D](#)). Second, the model can potentially predict an inversion of RT moments between compatibility conditions. Consider a scheme where the perceptual input of the irrelevant stimulus attribute p_{irrel} ⁴ is lower than that of the relevant attribute p_{rel} . This is plausible in the Simon task, because the location of the stimulus is not perceptually relevant, and should provide less evidence compared to the color. In compatible trials, the constant drift rate would be partly determined by p_{irrel} . The shrinking of attention in incompatible trials would cause the drift rate to converge toward p_{rel} and become progressively stronger compared to that of compatible trials. This scheme leads to a reduction of RT variability for incompatible trials and thus to an inconsistent RT moment ordering between compatibility conditions. For the sake of completeness, we also fit an alternative version of the DSTP with no late selection in compatible trials.

4. Model fitting

4.1. Method

Time-varying diffusion models were tested against group data from the previous Eriksen and Simon experiments. Probabilities and RT quantiles (.1, .3, .5, .7, .9) for correct and error responses for each condition were averaged across subjects. The .1 quantile represents the distribution’s leading edge, and the .9 quantile represents its tail. Only the median quantile (central tendency) was used for 35%, 45%, 60%, and 80% chroma levels in the compatible condition because the number of error responses was low (see [Table 1](#)). The SSP, DSTP, and the two alternative model versions were simulated as random walks (see [Section 2](#)), and were fitted to data using a SIMPLEX routine that minimizes the G^2 likelihood ratio statistic ([Ratcliff & Smith, 2004](#)):

$$G^2 = 2 \left(\sum_{i=1}^{12} n_i \sum_{j=1}^X p_{ij} \log \frac{p_{ij}}{\pi_{ij}} \right)$$

The outer summation i extends over the six chroma levels within each of the two compatibility conditions. n_i is the averaged number of valid trials per condition. The variable X represents the number of bins bounded by RT quantiles for each distribution pair of correct and error responses. We set $X = 8$ (6 bins for correct responses and 2 bins for errors) for 35–80% chroma levels in the compatible condition and $X = 12$ otherwise. p_{ij} and π_{ij} are respectively the observed and predicted proportions of responses in bin j of condition i . In this way, the model has to account for RT distribution shapes and choice probabilities simultaneously. 80,000 trials were simulated for each condition and each SIMPLEX iteration. In line with previous work (e.g., [Hübner et al., 2010](#); [Smith & Ratcliff, 2009](#)), the G^2 statistic was considered as a measure of relative fit quality, and was completed by a BIC that penalizes models according to their number of free parameters m :

$$\text{BIC} = G^2 + m \log \sum_{i=1}^{12} n_i$$

The goodness of fit of the models can also be appreciated graphically in [Figs. 8 and 9](#), where observed and predicted quantile probability functions (QPFs; [Ratcliff, 2001](#)) are superimposed. QPFs are constructed by aligning RT quantiles (y -axis) on the corresponding response type proportion (x -axis). For example, if the probability of a correct response in a given experimental condition is $p(c)$, the RT distributions of correct and error responses will be respectively aligned on $p(c)$ and $1 - p(c)$. Observed QPFs from the previous experiments reveal that color desaturation increases the

⁴ Note the subscripts denoting targets and flankers (tar, fl) are now replaced by relevant and irrelevant stimulus attributes ($rel, irrel$) to allow for more generality.

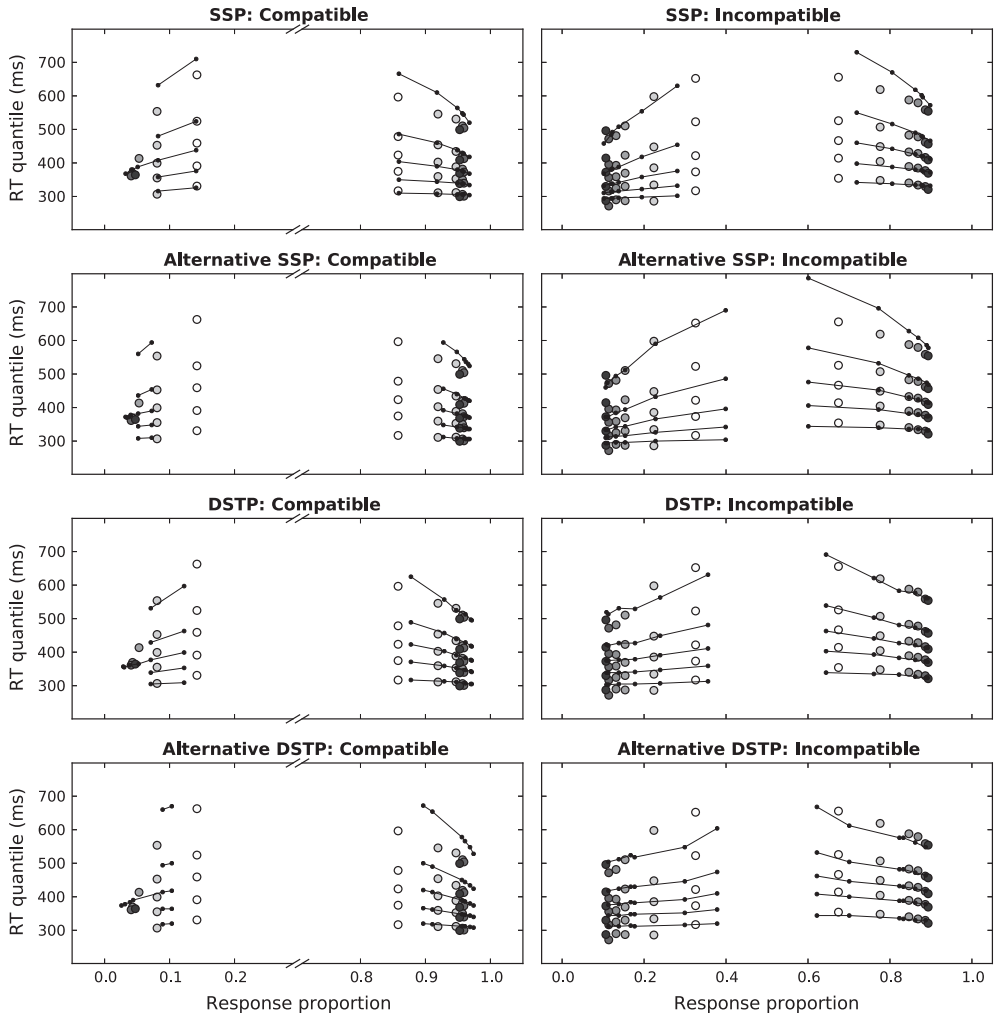


Fig. 8. Quantile probability functions (QPFs) for compatible and incompatible conditions across chroma levels, Experiment 1. Behavioral data are shown with circular markers. Chroma levels are symbolized by gray shading. Solid lines are fits of the different time-varying diffusion models illustrated in different panels.

mean, SD, and skew of RT distributions, as classically observed when stimulus discriminability is manipulated (e.g., Ratcliff & Smith, 2004). The effect of S–R compatibility is also consistent with previous work (e.g., White, Ratcliff, et al., 2011), with faster errors than correct responses for incompatible trials only. In Appendix E, we provide an alternative representation of the data and model predictions as CAFs. Although the models were not fitted to the CAFs (but rather to the QPFs), those functions represent the evolution of accuracy as a function of RT quantile, and may provide useful qualitative information about potential misfits that may not be directly obvious by inspecting QPFs (for a similar approach, see White, Brown, et al., 2011).

The architecture of the SSP for the Simon task is identical to that of the Eriksen, except that the Gaussian spotlight centers on the relevant color feature of the stimulus. The color region is defined as 1 unit wide, and the remaining attention is allocated to the irrelevant spatial feature. Alternative versions of the SSP and DSTP are respectively characterized by a lack of attentional shrinking and a lack of late stimulus selection in compatible trials only.

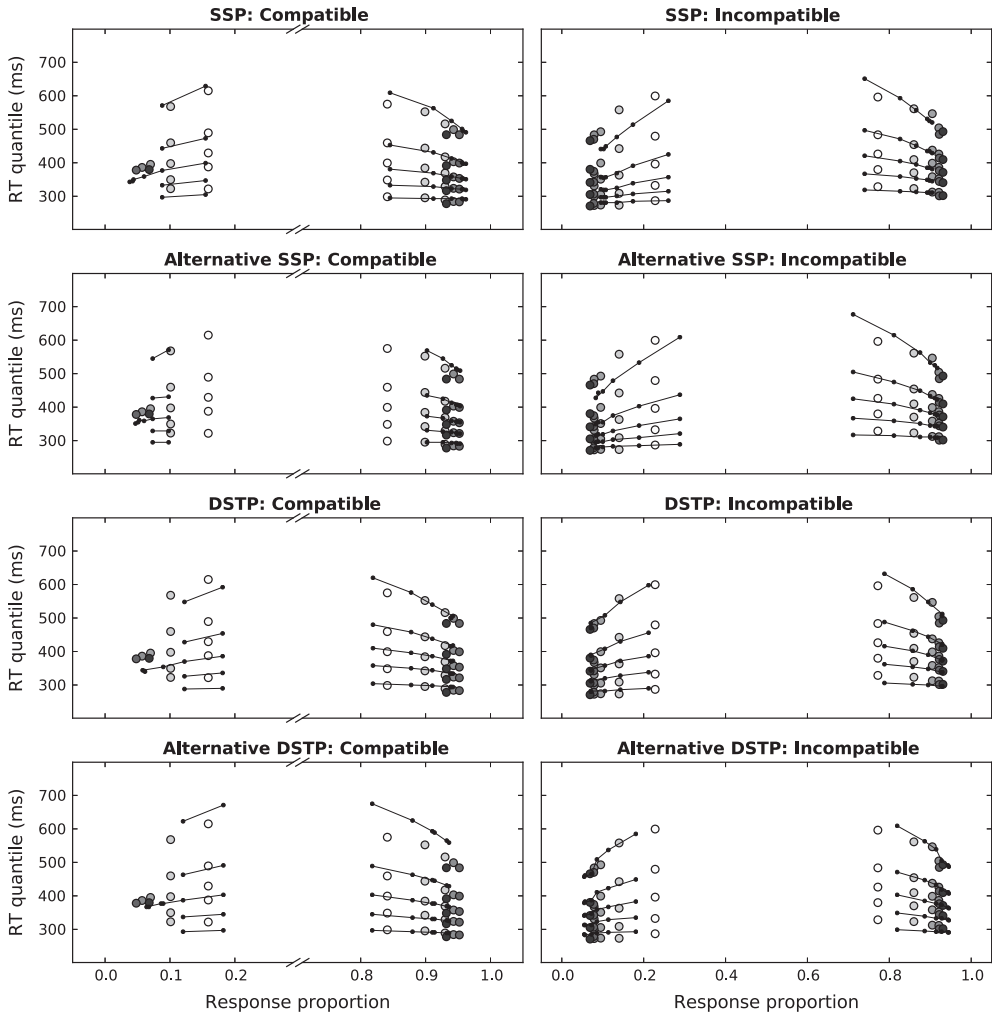


Fig. 9. Quantile probability functions (QPFs) for compatible and incompatible conditions across chroma levels, Experiment 2. Behavioral data are shown with circular markers. Chroma levels are symbolized by gray shading. Solid lines are fits of the different time-varying diffusion models illustrated in different panels.

80,000 Trials per experimental condition and fit cycle were simulated. Different starting points were used to ensure that the SIMPLEX gradient descent does not reach a local minimum in the parameter space. No parameter was allowed to vary between compatibility conditions. Boundary separations were fixed across chroma levels due to the randomized design of the experiments. The non-decision time T_{er} and the drift rate for the response selection process in phase two u_{rs2} in the DSTP were also fixed since variations of these parameters do not necessarily lead to Wagenmakers–Brown’s law (see Wagenmakers & Brown, 2007 and Section 2.2). To account for the experimental manipulation, parameters related to the perception/identification of the relevant stimulus attribute (p_{rel}^5 in the SSP, μ_{rel} and μ_{ss} in the DSTP) were allowed to vary across chroma levels. A model variant

⁵ p_{rel} was further constrained to be equal to p_{irrel} at 80% chroma in the Eriksen task. At this color saturation level, target and flankers are similar, and provide the same quantity of sensory evidence.

of the SSP in which the spotlight shrinking rate r_d was allowed to vary was also fitted to data. Because r_d variations were very small and had a negligible impact on the fit quality (see Appendix F), r_d was fixed.

4.2. Model fits: Eriksen task

Best-fitting parameters and fit statistics of the models are summarized in Table 4. Parameters are evolving as expected across chroma levels. The performance of the models can be graphically appreciated in Fig. 8. Original versions of the SSP and DSTP capture the main patterns of the data. However, the SSP overestimates the skew (i.e., tail quantile) of RT distributions for correct responses as chroma lessens. By contrast, the DSTP captures fairly well the variations of RT distribution shape for correct and error responses across conditions, although predicted errors are too fast for the lowest chroma

Table 4
Model parameters and fit statistics for Experiment 1 (Eriksen task).

	Chroma levels (%)					
	80	60	45	35	25	15
<i>SSP</i>						
<i>a</i>	0.100					
<i>Ter</i>	0.256					
p_{rel}	0.340	0.306	0.301	0.279	0.220	0.155
p_{irrel}	0.340					
sd_a	1.114					
r_d	0.013					
$G^2 = 167.9$						
$BIC = 245.4$						
<i>Alternative SSP</i>						
<i>a</i>	0.102					
<i>Ter</i>	0.256					
p_{rel}	0.333	0.324	0.290	0.266	0.192	0.091
p_{irrel}	0.333					
sd_a	1.148					
r_d	0.014					
$G^2 = 182.1$						
$BIC = 259.5$						
<i>DSTP</i>						
<i>a</i>	0.113					
<i>c</i>	0.178					
<i>Ter</i>	0.241					
μ_{rel}	0.201	0.196	0.170	0.134	0.104	0.065
μ_{irrel}	0.086					
μ_{ss}	0.350	0.353	0.340	0.339	0.290	0.198
μ_{rs2}	1.833					
$G^2 = 62.1$						
$BIC = 193.7$						
<i>Alternative DSTP</i>						
<i>a</i>	0.107					
<i>c</i>	0.192					
<i>Ter</i>	0.255					
μ_{rel}	0.236	0.215	0.189	0.180	0.103	0.089
μ_{irrel}	0.115					
μ_{ss}	0.405	0.383	0.373	0.370	0.327	0.242
μ_{rs2}	1.851					
$G^2 = 165.1$						
$BIC = 296.7$						

Note. *a* = boundary separation for the response selection process; *Ter* = non-decision time; p_{rel} = perceptual input of the relevant stimulus attribute; p_{irrel} = perceptual input of the irrelevant stimulus attribute; sd_a = spotlight width; r_d = spotlight shrinking rate; μ_{rel} = component rate for the relevant stimulus attribute; μ_{irrel} = component rate for the irrelevant stimulus attribute; *c* = boundary separation for the target identification process; μ_{ss} = drift rate for the target identification process; μ_{rs2} = drift rate for response selection in phase 2.

level in the compatible condition (see [Appendix E](#), for additional model analyses based on CAFs). Consequently, the DSTP provides a superior goodness-of-fit compared to the SSP, quantified by lower G^2 values. The BIC also favors the DSTP, despite a higher flexibility (17 free parameters for the DSTP against 10 for the SSP).

Focusing on mean RT for correct responses reveals an interesting phenomenon. [Fig. 10](#) shows the predicted Wagenmakers–Brown’s laws from best-fitting models. As can be seen, the compatibility effect predicted by the SSP increases monotonically from 41 ms (80% chroma) to 54 ms (15% chroma), and the compatibility factor affects both the slope and the intercept of Wagenmakers–Brown’s law, consistent with our initial simulation of the model (see [Section 2.1](#)). However, the compatibility effect predicted by the DSTP remains relatively constant across chroma levels (80% chroma: 41 ms; 60%: 40 ms; 45%: 40 ms; 35%: 38 ms; 25%: 41 ms; 15%: 42 ms), and the compatibility factor lowers the intercept of Wagenmakers–Brown’s law, without affecting its slope. While this pattern is in perfect agreement with the empirical data, it is at odds with our initial simulation (see [Section 2.2](#)). There are two main differences between the set of parameters used in [Section 2.2](#) and that obtained from fitting the DSTP to data. (i) Our initial simulation of the DSTP was based on fits reported by [White, Ratcliff, et al. \(2011, Experiment 1\)](#) for a standard Eriksen task. Those fits indicate a very high drift rate for the stimulus identification process μ_{ss} (1.045) and a lower drift rate for the response selection process in phase two μ_{rs2} (0.414). However, the fits of the DSTP reported by [Hübner and collaborators \(Hübner & Töbel, 2012; Hübner et al., 2010\)](#) consistently show the reversed pattern,⁶ i.e. lower drift rates for μ_{ss} (range 0.2913–0.5343) compared to μ_{rs2} (range 1.016–1.9799). This indicates a tradeoff between the two parameters, and the model seems to balance the first and second phases of response selection (i.e., slower first phase requires faster second phase). Our fits fall in the range of values reported by [Hübner and collaborators](#). A lower drift rate for μ_{ss} compared to μ_{rs2} appears more plausible because stimulus identification (μ_{ss}) is theoretically constrained by the physical properties of the stimulus while μ_{rs2} is not: μ_{rs2} is driven by the selected target (red or blue), and incorporates a strong manifestation of top-down control. (ii) In our initial simulation of the DSTP, μ_{ss} decreased from 1.045 to 0.445 while best-fitting values have a much smaller range (from 0.333 at 80% chroma to 0.198 at 15% chroma; see [Table 4](#)). Because the compatibility factor only affects the first phase of response selection, a higher variation of μ_{ss} leaves more time for interference to increase before the second phase of response selection arises. The combined influence of (i) and (ii) explains the different predictions of the DSTP.

Alternative versions of the SSP and the DSTP produced worse fit statistics compared to original ones. Removing the late selection process of the DSTP in the compatible condition strongly increases the skew of predicted RT distributions for correct responses. The alternative SSP underestimates the range of accuracy values in the compatible condition. The lack of attentional shrinking makes the drift rate partly determined by the flankers which remain at maximal intensity. This property prevents the model from capturing the augmentation of error rate when chroma decreases.

4.3. Model fits: Simon task

Parameters that yielded the best fit to the Simon data evolve across chroma levels in a similar manner compared to those of the Eriksen ([Table 5](#)). As shown in [Fig. 9](#), several misfits are apparent. Most importantly, the models do not capture the inversion of RT moments between compatibility conditions for each chroma level ([Fig. 10](#)). Obviously, the attempt to capture this statistical peculiarity has created collateral distortions. The DSTP predicts a Simon effect on mean RT that is too small and errors that are too fast in the compatible condition. The SSP predicts errors that are too fast in all conditions. An inspection of the CAFs (see [Appendix E](#)) reveals a surprising failure of the DSTP to explain accuracy dynamics across conditions. CAF shapes are better predicted by the SSP.

Alternative model versions are penalized by the same problems highlighted in the Eriksen task, and do not provide a better fit quality compared to original versions (indicated by higher G^2 and BIC statistics). The alternative DSTP overestimates the skew of RT distributions in the compatible condition

⁶ [Hübner and Töbel \(2012, Experiment 1\)](#) replicated [White, Ratcliff, et al.’s \(2011, Experiment 1\)](#) results and reported $\mu_{ss} = 0.4437$ and $\mu_{rs2} = 1.9799$.

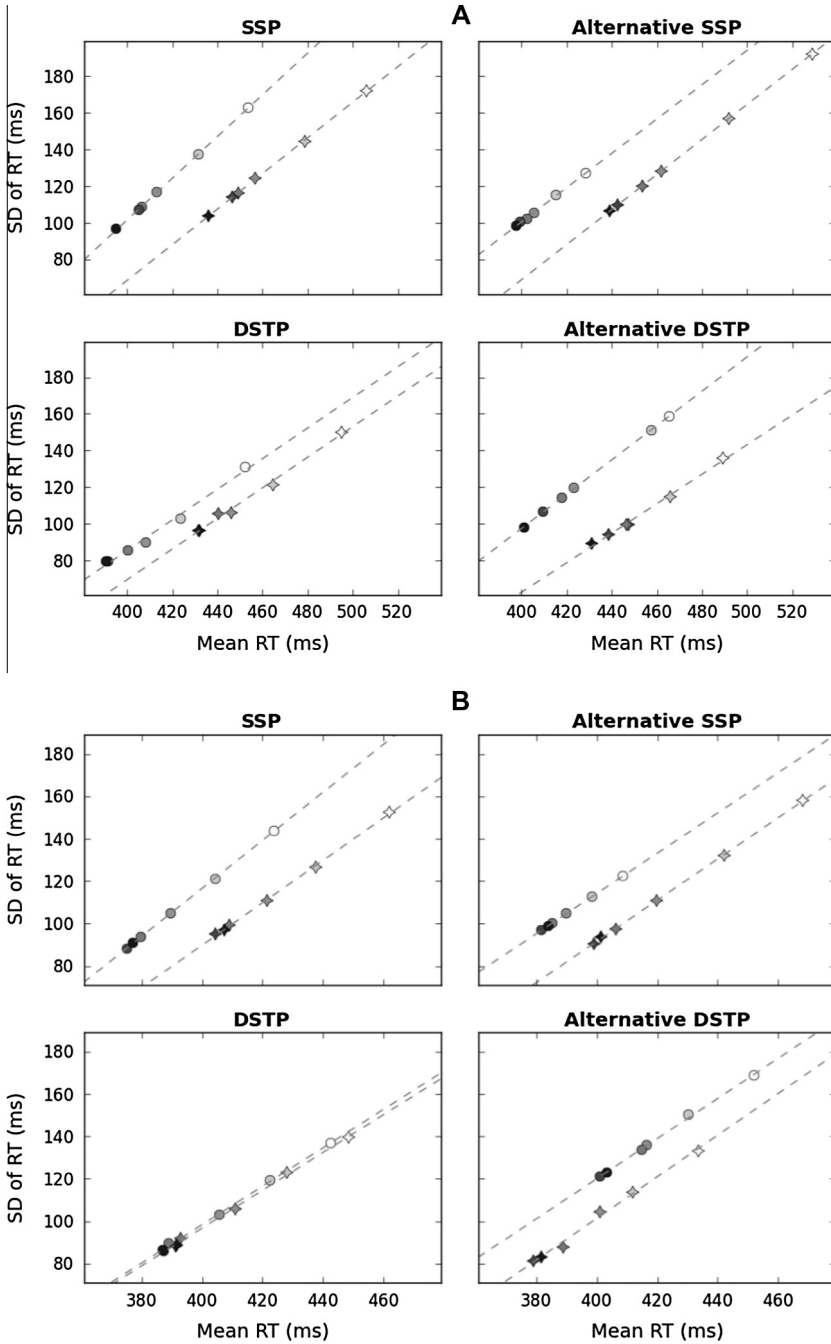


Fig. 10. Wagenmakers–Brown’s laws predicted by best-fitting diffusion models to the data of Experiment 1 (upper panel A) and Experiment 2 (lower panel B). Circles represent compatible conditions, stars represent incompatible conditions. Chroma levels are symbolized by gray shading. Also shown are lines of best fit independently for each compatibility mapping (dashed lines).

Table 5
Model parameters and fit statistics for Experiment 2 (Simon task).

	Chroma levels (%)					
	80	60	45	35	25	15
<i>SSP</i>						
<i>a</i>	0.096					
<i>Ter</i>	0.245					
<i>p_{rel}</i>	0.338	0.347	0.330	0.288	0.237	0.167
<i>p_{irrel}</i>	0.295					
<i>sd_a</i>	1.099					
<i>r_d</i>	0.015					
$G^2 = 171.9$						
BIC = 258.8						
<i>Alternative SSP</i>						
<i>a</i>	0.094					
<i>Ter</i>	0.245					
<i>p_{rel}</i>	0.352	0.376	0.335	0.291	0.219	0.142
<i>p_{irrel}</i>	0.287					
<i>sd_a</i>	1.082					
<i>r_d</i>	0.015					
$G^2 = 183.1$						
BIC = 269.9						
<i>DSTP</i>						
<i>a</i>	0.114					
<i>c</i>	0.168					
<i>Ter</i>	0.220					
μ_{rel}	0.168	0.167	0.176	0.136	0.115	0.077
μ_{irrel}	0.011					
μ_{ss}	0.361	0.366	0.344	0.298	0.244	0.196
μ_{rs2}	1.871					
$G^2 = 129.2$						
BIC = 263.4						
<i>Alternative DSTP</i>						
<i>a</i>	0.102					
<i>c</i>	0.187					
<i>Ter</i>	0.235					
μ_{rel}	0.252	0.258	0.224	0.219	0.189	0.140
μ_{irrel}	0.007					
μ_{ss}	0.357	0.369	0.344	0.239	0.214	0.160
μ_{rs2}	1.921					
$G^2 = 226.1$						
BIC = 360.3						

Note. *a* = boundary separation for the response selection process; *Ter* = non-decision time; *p_{rel}* = perceptual input of the relevant stimulus attribute; *p_{irrel}* = perceptual input of the irrelevant stimulus attribute; *sd_a* = spotlight width; *r_d* = spotlight shrinking rate; μ_{rel} = component rate for the relevant stimulus attribute; μ_{irrel} = component rate for the irrelevant stimulus attribute; *c* = boundary separation for the target identification process; μ_{ss} = drift rate for the target identification process; μ_{rs2} = drift rate for response selection in phase 2.

and generates a reversed Simon effect on mean RT. The alternative SSP underestimates the range of accuracy values in the compatible condition, and the model predicts an inversion of RT moments between compatibility conditions only for the higher chroma values. In conclusion, none of the models evaluated are able to fit the Simon data.

5. General discussion

5.1. A common theoretical framework for different conflict tasks

On the basis of conceptual (e.g., Hommel, 2011; Kornblum et al., 1990) and statistical (Pratte et al., 2010; Speckman, Rouder, Morey, & Pratte, 2008) differences, it has long been argued that different

conflict tasks are likely to involve different components of processing. By concurrently investigating Piéron and Wagenmakers–Brown laws in Eriksen (Experiment 1) and Simon (Experiment 2) tasks, both at experimental and computational levels, we adopted a novel strategy to gain insight into decision-making in a conflicting environment. Our data identify strong processing similarities between tasks through three key findings. In both tasks, we found that (i) Piéron's law holds for each S–R compatibility condition. (ii) Compatibility and color saturation combine additively, as revealed by Bayesian hypothesis testing. (iii) Wagenmakers–Brown's law holds for color saturation, but is broken by the compatibility factor: the incompatible mapping lowers the intercept of the linear law while leaving its slope constant. Altogether, those results provide evidence for a common model base between Eriksen and Simon tasks (see [Burle, Spieser, Servant, & Hasbroucq, 2014](#), for electromyographic evidence supporting a similar conclusion). The recent findings of [Stafford et al. \(2011\)](#) suggest that the Stroop task may also belong to this common framework.⁷

The violation of Wagenmakers–Brown's law by the compatibility factor strongly deviates from optimal predictions of a standard DDM. As an alternative account, we explored a new generation of diffusion models that incorporate selective attention mechanisms. Simulations of the SSP and the DSTP showed that the violation of Wagenmakers–Brown's law by the compatibility factor was indeed predicted. The violation appears to be induced by an attention-driven improvement of the quality of sensory evidence over time (i.e., a progressive suppression of the irrelevant stimulus attribute influence), regardless whether attentional selectivity operates in a continuous or discrete manner. This dynamic results in a time-varying evidence accumulation process underlying decision-making under conflict.

5.2. A discrete improvement of attentional selectivity better explains processing in the Eriksen task

A further test of the DSTP and the SSP was carried out by fitting them to the RT distributions and accuracy data of our two experiments. So far, the models have only been tested against data from Eriksen tasks, and it has proven difficult to determine the superiority of one model over another due to substantial mimicry, despite different theoretical assumptions ([Hübner & Töbel, 2012](#); [White, Ratcliff, et al., 2011](#)). In this respect, the data from our Eriksen task appears particularly constraining and challenging: the models have to explain the variations of accuracy and the shape of RT distributions over the six color saturation levels and the two flanker compatibility conditions. Moreover, they must do this with fixed decision boundaries, only parameters related to the perception/identification of the target being free to vary across chroma levels. Comparative fits reveal a numerical advantage of the DSTP over the SSP. The DSTP fits all aspects of the Eriksen data reasonably well. The SSP has the problem that it overestimates the skew of RT distributions for correct responses as chroma decreases, whatever the compatibility mapping. This overestimation is more pronounced in the incompatible condition, and the model predicts a super-additive interaction between compatibility and chroma. The SSP also fails to capture qualitative patterns of the CAFs across conditions. These failures could be due to any component of the model. In particular, we treated non-decision time, moment-to-moment noise and between-trial variability in drift rate as fixed parameters in the fits reported here, but those parameters could be plausibly affected by chroma. Relaxing any of these constraints may virtually improve the fit quality of the SSP. Alternatively, the failures of the model may be rooted in its general single-stage assumption. Because stimulus identification and response selection are embodied in a single decision process, the drift rate is always constrained by the physical properties of the stimulus, even late in the course of processing (the drift rate converges toward the perceptual input of the target). By contrast, the DSTP assumes that stimulus identification and response selection are two separate and parallel processes. When a stimulus is identified, response selection takes another drift rate (μ_{rs2}) unconstrained by the physical properties of the stimulus, and driven exclusively by the selected stimulus. This second and more efficient process allows the model to capture the shape of observed RT distributions for correct responses across conditions.⁸ As a consequence,

⁷ [Stafford et al. \(2011\)](#) did not explore Wagenmakers–Brown's law. We hypothesize that finding (iii) also holds in their Stroop data set.

⁸ Indeed, removing the late selection stage strongly increases the skew of predicted RT distributions for correct responses.

the DSTP explains the pattern of results regarding psychological laws and additive effects between experimental factors. The latter finding deserves consideration.

Additive effects between a S–R compatibility factor and variables that affect perceptual processing have consistently been observed (for reviews, see Sanders, 1980, 1990). S–R compatibility effects have been shown to combine additively with target duration (Simon & Berbaum, 1990), target eccentricity (Hommel, 1993, Experiment 1), and target quality (e.g., Acosta & Simon, 1976; Everett, Hochhaus, & Brown, 1985; Frowein & Sanders, 1978; Sanders, 1977; Shwartz, Pomerantz, & Egeth, 1977; Simon, 1982; Simon & Pouraghabagher, 1978; Stoffels, Van der Molen, & Keuss, 1985; van Duren & Sanders, 1988; but see Hommel, 1993, Experiments 2–5; Stanovich & Pachella, 1977). Target quality has been manipulated along various dimensions such as signal-background luminance contrast, sound bursts intensity levels, or visual noise. Hence, our results and those of Stafford et al. (2011) cannot be due to a peculiarity of color saturation.⁹

Simulations of the DSTP performed in the present work show that the model is able to generate different outcomes (additivity/super-additivity between color saturation and compatibility, linear/curvilinear relationship between the mean and SD of RT distributions) under seemingly plausible parametric variations. Moreover, they highlight a tradeoff between the first and second phase of response selection. The model appears so flexible that it may be difficult to falsify. However, the DSTP fails to explain the Simon data, showing that it is indeed falsifiable.

5.3. A failure to account for the Simon data

The results of our experiments suggest a common model framework for different conflict tasks. This finding appears problematic for the SSP because the model was specifically designed to account for spatial attention dynamics in the Eriksen task, although White, Ratcliff, et al. (2011) hypothesized that the spotlight component may also center on a more abstract attentional space. On the contrary, Hübner et al. (2010) formalized the DSTP in a sufficiently abstract way to “potentially serve as a framework for interpreting distributional effects in a large range of conflict paradigms” (p. 760). However, neither the DSTP nor the SSP explain processing in the Simon task, because the models are unable to predict an inversion of RT moments between compatibility conditions (i.e., the incompatible condition is associated with the largest mean and the *smallest* SD of RT) characteristic of the task (e.g., Burle et al., 2002; Pratte et al., 2010; Schwarz & Miller, 2012). This statistical peculiarity suggests an important parametric variation between Eriksen and Simon tasks. An inversion of RT moments may be generated by a rate of evidence accumulation that becomes progressively higher for the incompatible compared to the compatible condition. The reason for such a counter-intuitive scheme is unclear. We explored alternative versions of the SSP and the DSTP with a lack of attentional selection in compatible trials. Those models, motivated by a preservation of attentional effort principle (White, Ratcliff, et al., 2011), implement the above drift rate scheme. However, their fit quality in Eriksen and Simon tasks was numerically inferior compared to standard model versions. Therefore, the SSP and the DSTP appear incomplete.

Because the DSTP captures qualitative and quantitative aspects of the Eriksen data that the SSP cannot, its architecture may represent a better foundation for a unified framework. This conclusion should be tempered by two caveats. First, as mentioned in the previous section, relaxing some parameter constraints may lead to different model performances. Second, analysis of the CAFs in the Simon task reveals an important failure of the DSTP to account for accuracy dynamics across conditions, and the model appears to generate qualitatively wrong predictions. The SSP provides a superior fit. These observations deserve further investigations.

On the one hand, the need for at least one additional parameter seems to weaken the DSTP framework. The model components would sum to eight, which further increases the risk of parameter

⁹ Following the additive factor method (Sternberg, 1969), Sternberg (2013) recently argued that the additivity observed by Stafford et al. (2011) is consistent with two serially arranged stages, and a selective influence of color saturation and compatibility on those stages. However, such model does not predict Wagenmakers–Brown’s law but instead a linearity between the mean and variance of RT distributions (Townsend & Ashby, 1983), although the two alternatives may not always be easy to dissociate. Furthermore, we do not see how a serial stage model could predict the very fast errors triggered by the irrelevant stimulus attribute in incompatible trials (see CAFs, Appendix E).

tradeoffs. On the other hand, this cost may be necessary to capture the types of nuance that are hallmarks of decision-making in conflicting situations. Currently, the DSTP is a formal implementation of qualitative dual-route models (e.g., Kornblum et al., 1990) in the context of selective attention (Hübner et al., 2010). To explain the particular distributional data of the Simon task, Ridderinkhof (2002) refined dual-route models by hypothesizing a response-based inhibitory mechanism that takes time to build. Alternatively, Hommel (1993) proposed that irrelevant location-based activations spontaneously decay over time. Testing these hypotheses are beyond the scope of the present paper, but they should be considered in future extensions of the model. Importantly, any proposed theory should provide a principled account of the parametric variations observed between the different conflict tasks.

6. Conclusion

The present work introduced a novel strategy to provide additional insight into decision-making in conflicting situations. The concurrent investigation of Piéron and Wagenmakers–Brown’s laws in Eriksen and Simon tasks highlighted several important constraints for RT models and strongly suggested a common model framework for the two tasks. Recent extensions of the DDM that incorporate selective attention mechanisms represent a promising approach toward the achievement of this goal. Detailed analyses revealed that a discrete improvement of attentional selectivity, as implemented through the DSTP, better explains processing in the Eriksen task compared to a continuous-improvement SSP. However, the DSTP fails to capture a statistical peculiarity of the Simon data and requires further development. Our results set the groundwork for an integrative diffusion model of decision-making in conflicting environments.

Acknowledgments

The research was supported by a European Research Council under the European Community’s Seventh Framework Program (FP/2007–2013 Grant Agreement No. 241077) and by the CNRS. The authors wish to thank Corey White, Ronald Hübner, Scott Brown, Eric-Jan Wagenmakers and Thierry Hasbroucq for their helpful comments. We also thank Marcel Janssen for technical assistance with color calibration. Distributional data and Python codes of the models are available upon request.

Appendix A

In the SSP, the target’s region is defined as 1 unit wide, with the Gaussian distribution centered on it. Assuming that the Gaussian distribution is centered on 0, the quantity of attention allocated to the target is defined as:

$$a_{tar}(t) = \int_{-0.5}^{0.5} N(0, sd_a(t))$$

$$\text{where } sd_a(t) = sd_a(0) - r_t(t)$$

For a 3-items Eriksen display (1 target and 2 flankers), the quantity of attention allocated to the flankers is:

$$a_f(t) = 1 - a_{tar}(t)$$

The attentional shrinking mechanism stops when $sd_a(t) = 0.001$, to prevent the width of the Gaussian from becoming negative.

Appendix B

We provide an additional simulation of the DSTP, identical to that performed in Section 2.2, except that the drift rate of the response selection process in phase two (μ_{rs2}) increases from 0.414 to 0.49 as target intensity decreases. Fig. B.1 shows the resulting predictions. The relationship between the mean and SD of RT distributions for each compatibility condition is curvilinear.

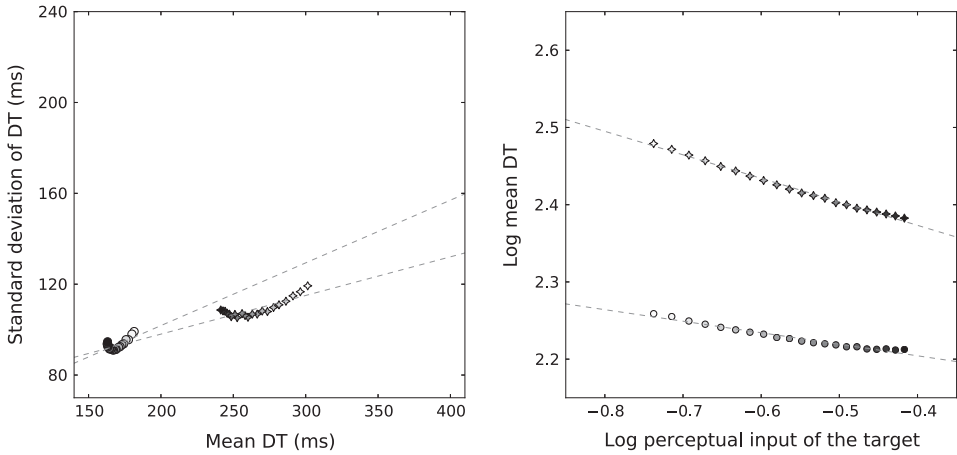


Fig. B.1. Predictions from the DSTP model for the Wagenmakers–Brown (left panel) and the log-transformed Piéron (right panel) laws, under the hypothesis that the drift rate for the response selection process in phase two increases when the perceptual intensity of the target decreases. Circles represent compatible conditions, stars represent incompatible conditions. Perceptual input levels of the target are symbolized by gray shading. Also shown are lines of best fit for each compatibility mapping (dashed lines). Left panel: SD of DT versus mean DT (both in milliseconds). Right panel: chronometric functions displayed in a log–log space.

Appendix C

In the main text, model selection was achieved using a BIC index and *p*-values were computed using Markov chain Monte Carlo (MCMC) sampling. The more parsimonious models were *random intercept models*. Although widely used in the literature, recent simulation studies have argued that random intercept models can have a catastrophically high Type I error rate, regardless of how *p*-values are computed from them (Barr et al., 2013; Schielzeth & Forstmeier, 2009). Therefore, the effect of RT mean and S–R compatibility may have been overestimated. Barr and colleagues recommended to include the maximal random effects structure justified by the design to avoid this problem. In our case, the maximal random effects structure would incorporate *by-subject random slopes and intercepts* with respect to our experimental factors. Unfortunately, those complex models have random correlation parameters, and deriving *p*-values from MCMC sampling is not possible in these particular cases. We thus

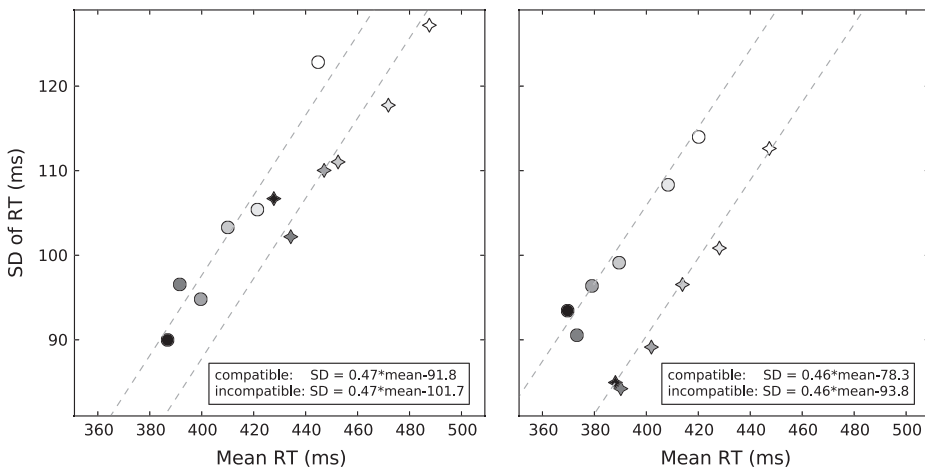


Fig. C.1. Standard deviation (SD) of response time (RT) versus mean RT across the 12 experimental conditions from Experiment 1 (left panel) and Experiment 2 (right panel). Data are averaged across subjects. Circles represent compatible conditions, stars represent incompatible conditions. Chroma levels are symbolized by gray shading. Dashed lines are constructed using the best-fitting parameters of a linear mixed effects model incorporating a full random effect structure. Their equation is provided in the inset.

performed likelihood-ratio tests to assess fixed effects (for technical details and validation of this statistical assessment, see Barr et al., 2013).

C.1. Eriksen task

We found effects of mean RT and S–R compatibility on the SD of RT distributions, $\chi^2(1) = 60.8$, $p < .001$ and $\chi^2(1) = 14.3$, $p < .001$ respectively (likelihood ratio tests). Moreover, the interaction between the two factors was not significant, $\chi^2(1) = 1.1$, $p = .3$. The best-fitting parameters for the fixed effects were similar to those obtained with the original method (see Fig. C.1, left panel).

C.2. Simon task

The results were similar to those obtained in the Eriksen task. We found effects of mean RT ($\chi^2(1) = 57.2$, $p < .001$), S–R compatibility ($\chi^2(1) = 24.8$, $p < .001$), and no interaction between the 2 factors ($\chi^2(1) = 0$). The best-fitting parameters for the fixed effects were again similar to those obtained with the original method (Fig. C.1, right panel).

Appendix D

Alternative versions of the SSP and DSTP are respectively characterized by a lack of attentional shrinking and a lack of late stimulus selection in compatible trials. As the original versions, the models predict Piéron and Wagenmakers–Brown laws for each compatibility condition (Fig. D.1). Note a peculiarity of the alternative SSP is its prediction that the compatibility factor lowers the intercept of Wagenmakers–Brown's law but does not affect its slope.

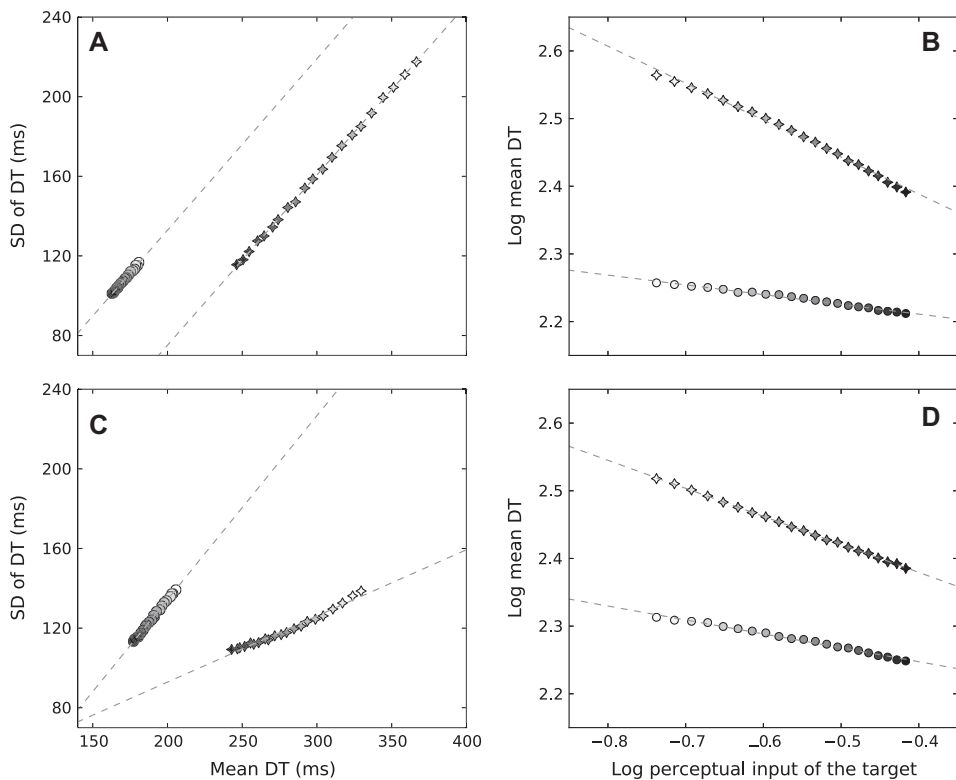


Fig. D.1. Predictions from the alternative SSP (A, B) and alternative DSTP (C, D) diffusion models for the Wagenmakers–Brown (left panels) and the log-transformed Piéron (right panels) laws when the perceptual intensity of the target decreases while that of the irrelevant stimulus attribute remains maximal. Circles represent compatible conditions, stars represent incompatible conditions. Perceptual input levels of the target are symbolized by gray shading. Also shown are lines of best fit for each compatibility mapping (dashed lines). Left panels: SD of DT versus mean DT (both in milliseconds). Right panels: chronometric functions displayed in a log–log space.

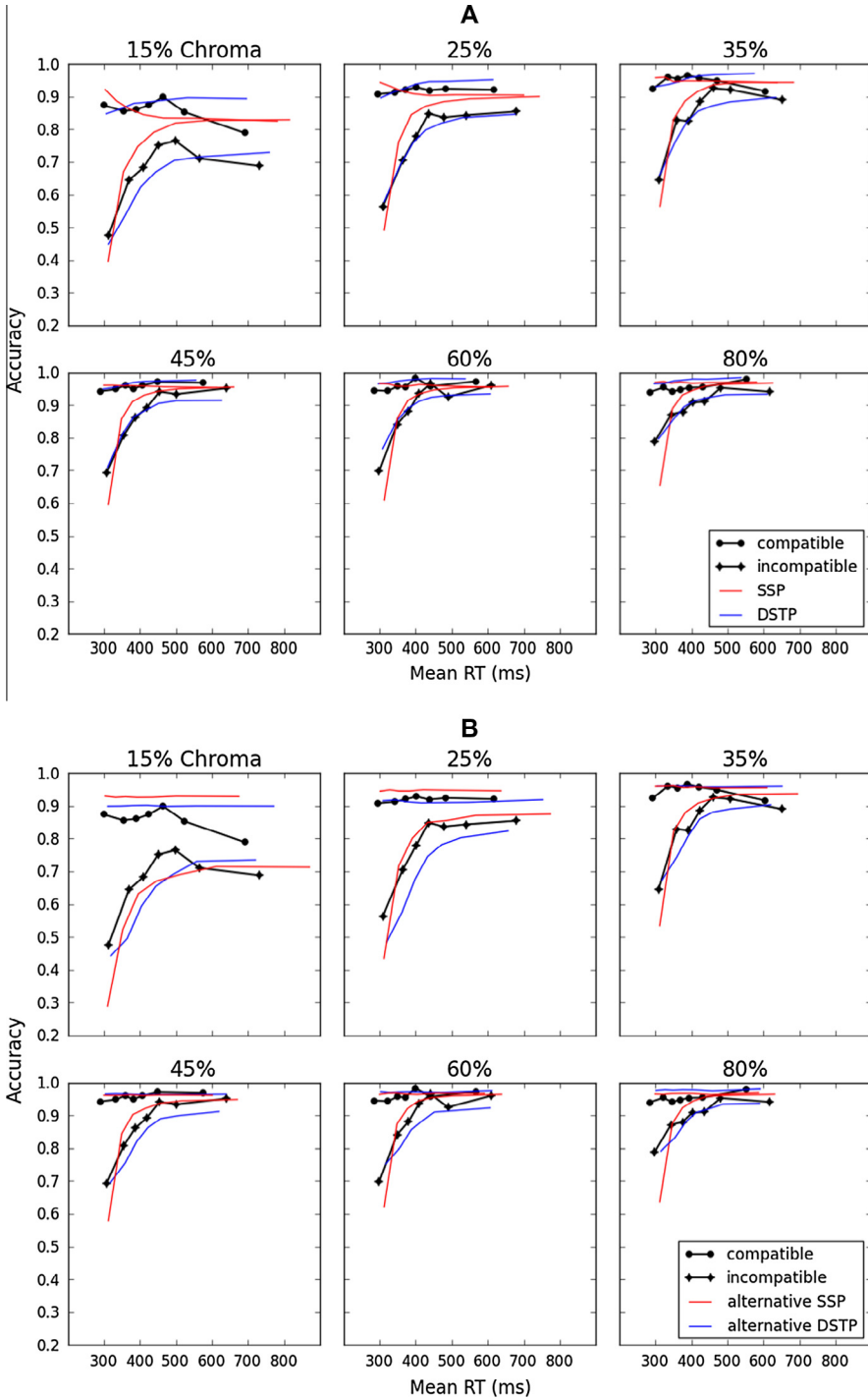


Fig. E.1. Observed and predicted conditional accuracy functions across experimental conditions, Experiment 1.

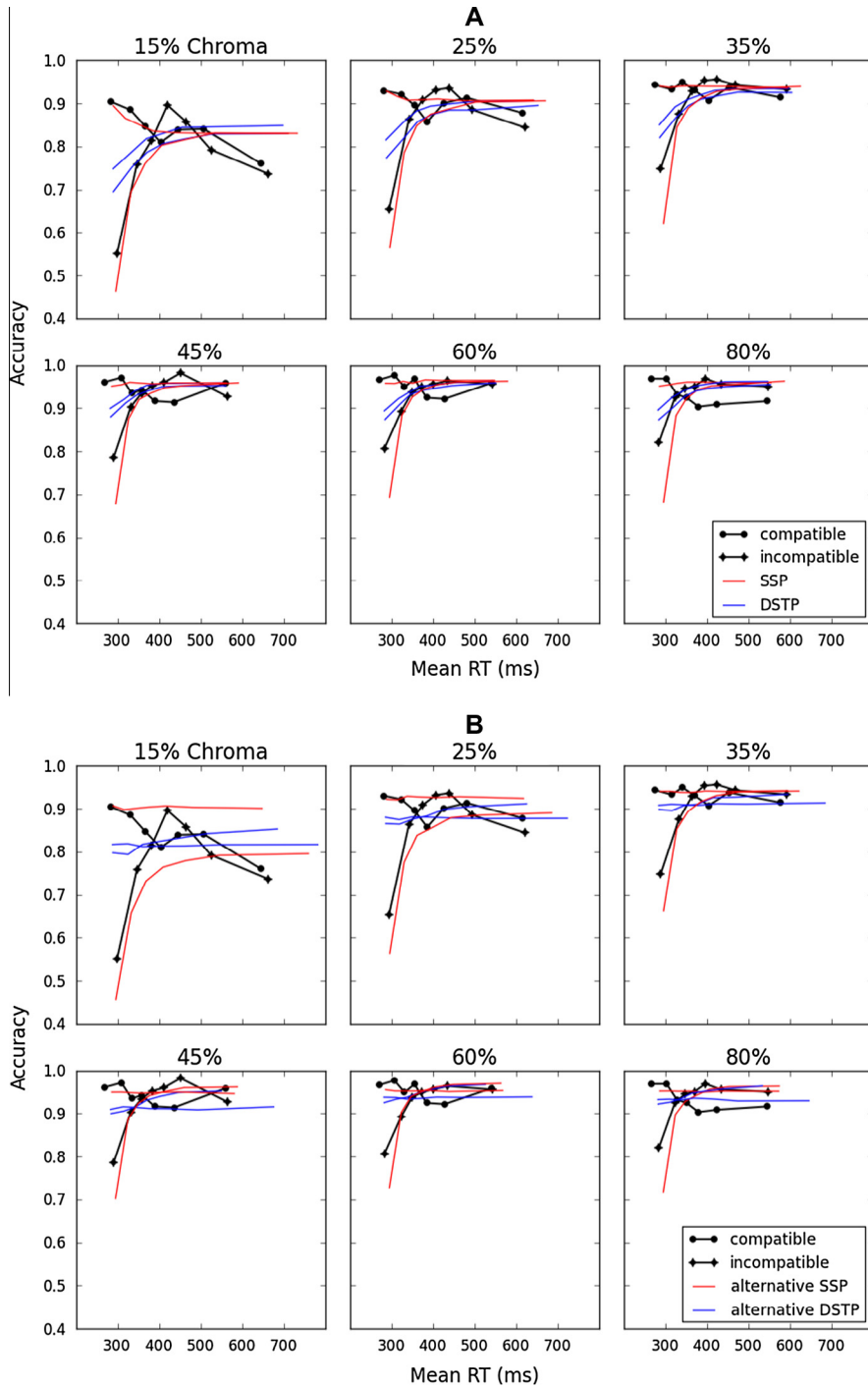


Fig. E.2. Observed and predicted conditional accuracy functions across experimental conditions, Experiment 2.

Table F.1
Model parameters and fit statistics for Experiment 1 (Eriksen task).

	Chroma levels (%)					
	80	60	45	35	25	15
SSP						
a	0.100					
Ter	0.256					
p_{rel}	0.340	0.337	0.302	0.279	0.223	0.143
p_{irrel}	0.340					
sd_a	1.114					
r_d	0.0151	0.0128	0.0140	0.0131	0.0134	0.0138
$G^2 = 162.1$						
BIC = 261.1						

Note. a = boundary separation for the response selection process; Ter = non-decision time; p_{rel} = perceptual input of the relevant stimulus attribute; p_{irrel} = perceptual input of the irrelevant stimulus attribute; sd_a = spotlight width; r_d = spotlight shrinking rate.

Appendix E

Fig. E.1 (Eriksen task) and Fig. E.2 (Simon task) display an alternative representation of the data and predictions from best-fitting models as CAFs. For each experimental condition, RT were rank-ordered and partitioned in 7 bins of equal size. The CAFs were constructed by plotting the proportion of correct responses in each bin against the average RT in the corresponding bin. Seven bins were chosen to provide a detailed representation of the dynamic of accuracy. The observed CAFs show the typical dip of accuracy for faster RT bins in the incompatible condition (e.g., Gratton et al., 1988). The proportion of these fast errors increases as chroma decreases.

The following analysis of predictions from best-fitting models is qualitative, because the models were not fitted to the CAFs (but rather to the QPFs). In the Eriksen task, for each chroma level, the SSP overestimates the proportion of very fast errors in the incompatible condition and predicts an improvement of accuracy that is too fast. The DSTP better captures the shape of the CAFs across conditions. These observations are consistent with the recent findings of Hübner and Töbel (2012). The SSP also predicts that responses in compatible and incompatible conditions converge toward the same accuracy rate as RT increases, because their drift rates converge toward the same value (equal to the perceptual input of the relevant stimulus attribute p_{rel}) as attention shrinks. However, the observed CAFs for 15% and 25% chroma levels show that accuracy is higher for compatible than incompatible trials, even for the slowest responses. The DSTP and the alternative version of the SSP do not predict such a convergence of the CAFs for 15% and 25% chroma levels, and better capture this aspect of the data.

Contrary to the Eriksen task, the SSP provides a superior fit than the DSTP in all conditions of the Simon. The DSTP has the serious problem that it predicts similar CAF dynamics for compatible and incompatible mappings. Both models, however, are unable to explain a peculiarity of the data: for each chroma level, the CAFs systematically cross in the medium range of RT latencies, leading to the counter-intuitive finding that accuracy rate becomes progressively higher in the incompatible compared to the compatible condition. The alternative SSP is able to predict this pattern, but only for the higher chroma levels. We suspect that the inversion of RT moments and the crossing of the CAFs between compatibility conditions are underpinned by the same mechanism, namely a rate of evidence accumulation that becomes progressively higher in the incompatible compared to the compatible condition.

Appendix F

Table F.1 shows the best fitting parameters and fit statistics of the SSP to the Eriksen data when the perceptual input of the relevant stimulus attribute p_{rel} and the spotlight shrinking rate parameter r_d are allowed to vary across chroma levels. Analysis of r_d variations reveals that it is slightly more difficult to focus on a target that has a lower color saturation compared to the flankers. However, the variations are very small, and do not significantly improve the fit quality.

References

- Acosta, E., & Simon, J. R. (1976). The effect of irrelevant information on the stages of processing. *Journal of Motor Behavior*, 8, 181–187. <http://dx.doi.org/10.1080/00222895.1976.10735070>.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <http://dx.doi.org/10.1016/j.jml.2007.12.005>.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D. M., Maechler, M., & Bolker, B. (2012). *lmer4: Linear mixed-effects models using Eigen and S4 classes*. R package (version 0.999999-0).

- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–765.
- Bonnet, C. (1992). Psychophysical scaling within an information processing approach? *Behavioral and Brain Sciences*, *15*, 560–561.
- Burle, B., Spieser, L., Servant, M., & Hasbroucq, T. (2014). Distributional reaction time properties in the Eriksen task: Marked differences or hidden similarities with the Simon task? *Psychonomic Bulletin & Review*. <http://dx.doi.org/10.3758/s13423-013-0561-6>.
- Burle, B., Possamai, C. A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research Psychologische Forschung*, *66*, 324–336. <http://dx.doi.org/10.1007/s00426-002-0105-6>.
- Chocholle, R. (1940). Variation des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *Année Psychologique*, *41*, 65–124. <http://dx.doi.org/10.3406/psy.1940.5877>.
- Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.
- Commission Internationale de l'Eclairage (1976). ISO 11664-4:2008(E)/CIE S 014-4/E:2007: Joint ISO/CIE Standard: Colorimetry—Part 4: CIE 1976 L*a*b* Colour Space.
- Davranche, K., Hall, B., & McMorris, T. (2009). Effect of acute exercise on cognitive control required during an Eriksen flanker task. *Journal of Sport and Exercise Psychology*, *31*, 628–639.
- Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*, 304–322. [http://dx.doi.org/10.1016/S0022-2496\(03\)00003-8](http://dx.doi.org/10.1016/S0022-2496(03)00003-8).
- 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. <http://dx.doi.org/10.3758/BF03203267>.
- 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.
- Everett, B. L., Hochhaus, L., & Brown, J. R. (1985). Letter-naming as a function of intensity, degradation, S–R compatibility, and practice. *Perception & Psychophysics*, *37*, 467–470.
- Frowein, H. W., & Sanders, A. F. (1978). Effects of visual stimulus degradation, S–R compatibility, and foreperiod duration on choice reaction time and movement time. *Bulletin of the Psychonomic Society*, *12*(2), 106–108. <http://dx.doi.org/10.3758/BF03329641>.
- Gratton, G., Coles, M. G., 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.
- Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.
- Hasbroucq, T., Possamai, C. A., Bonnet, M., & Vidal, F. (1999). Effect of the irrelevant location of the response signal on choice reaction time: An electromyographic study in humans. *Psychophysiology*, *36*, 522–526.
- Hommel, B. (1993). The relationship between stimulus processing and response selection in the Simon task: Evidence for a temporal overlap. *Psychological Research Psychologische Forschung*, *55*, 280–290. <http://dx.doi.org/10.1007/BF00419688>.
- Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, *136*, 189–202.
- Hübner, R., Steinhäuser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, *117*, 759–784.
- Hübner, R., & Töbel, L. (2012). Does attentional selectivity in the flanker task improve discretely or gradually? *Frontiers in Psychology*, *3*, 1–11. <http://dx.doi.org/10.3389/fpsyg.2012.00434>.
- Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility – A model and taxonomy. *Psychological Review*, *97*, 253–270.
- Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York: Wiley.
- Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics*, *72*, 246–273.
- Luce, R. D. (1986). *Response times*. New York: Oxford University Press.
- Morey, R. D., & Rouder, J. N. (2012). *BayesFactor*. R package (version 0.9.5).
- Nelder, J. A., & Mead, R. (1965). A Simplex method for function minimization. *The Computer Journal*, *7*, 308–313. <http://dx.doi.org/10.1093/comjnl/7.4.308>.
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*, 376–404.
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. <http://dx.doi.org/10.1016/j.jneumeth.2006.11.017>.
- Piéron, H. (1913). Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices. *Année Psychologique*, *22*, 17–96. <http://dx.doi.org/10.3406/psy.1913.4294>.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Pins, D., & Bonnet, C. (1996). On the relation between stimulus intensity and processing time: Pieron's law and choice reaction time. *Perception & Psychophysics*, *58*, 390–400.
- Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*, *72*, 2013–2025.
- 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. (2001). Diffusion and random walks models. *International encyclopedia of the social and behavioral sciences* (Vol. 6, pp. 3668–3673). Oxford, England: Elsevier.

- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922. <http://dx.doi.org/10.1162/neco.2008.12-06-420>.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356. <http://dx.doi.org/10.1111/1467-9280.00067>.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111, 333–367.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9, 438–481.
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461, 263–266.
- Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action. Attention and performance* (pp. 494–519). Oxford: Oxford University Press.
- Ridderinkhof, K. R., Scheres, A., Oosterlaan, J., & Sergeant, J. A. (2005). Delta plots in the study of individual differences: New tools reveal response inhibition deficits in AD/Hd that are eliminated by methylphenidate treatment. *Journal of Abnormal Psychology*, 114, 197–215.
- Rösler, F., & Finger, T. (1993). A psychophysiological analysis of response-channel activation and outcome states in Eriksen's noise-compatibility task. *Psychological Research Psychologische Forschung*, 55, 20–28.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356–374. <http://dx.doi.org/10.1016/j.jmp.2012.08.001>.
- Sanders, A. F. (1990). Some issues and trends in the debate on discrete vs. continuous processing of information. *Acta Psychologica*, 74, 123–167.
- Sanders, A. F. (1977). Structural and functional aspects of the reaction process. In S. Dornic (Ed.), *Attention and performance 6* (pp. 3–25). Hillsdale, NJ: Erlbaum.
- Sanders, A. F. (1980). Stage analysis of reaction processes. In G. E. Stelmach & J. Q. Requin (Eds.), *Tutorials in motor behavior* (pp. 331–354). Amsterdam: North-Holland.
- Schieleth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, 20, 416–420. <http://dx.doi.org/10.1093/beheco/arn145>.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Schwarz, W., & Miller, J. (2012). Response time models of delta plots with negative-going slopes. *Psychonomic Bulletin & Review*, 19, 555–574. <http://dx.doi.org/10.3758/s13423-012-0254-6>.
- Shwartz, S. P., Pomerantz, J. R., & Egeth, H. E. (1977). State and process limitations in information processing: An additive factors analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 402–410.
- Simon, J. R. (1982). Effect of an auditory stimulus on the processing of a visual stimulus under single- and dual-tasks conditions. *Acta Psychologica*, 51, 61–73. [http://dx.doi.org/10.1016/0001-6918\(82\)90019-1](http://dx.doi.org/10.1016/0001-6918(82)90019-1).
- Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The 'Stroop effect' vs. the 'Simon effect'. *Acta Psychologica*, 73, 159–170.
- Simon, J. R., & Pouraghabagher, A. R. (1978). The effect of aging on the stages of processing in a choice reaction time task. *Journal of Gerontology*, 33, 553–561.
- Simon, J. R., & Small, A. M. Jr., (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, 53, 433–435.
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, 116, 283–317.
- Speckman, P. L., Rouder, J. N., Morey, R. D., & Pratte, M. S. (2008). Delta plots and coherent distribution ordering. *American Statistician*, 62, 262–266.
- Stafford, T., Ingram, L., & Gurney, K. N. (2011). Pieron's Law holds during Stroop conflict: Insights into the architecture of decision making. *Cognitive Science*, 35, 1553–1566. <http://dx.doi.org/10.1111/j.1551-6709.2011.01195.x>.
- Stanovich, K. E., & Pachella, R. G. (1977). Encoding, stimulus-response compatibility, and stages of processing. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 411–421.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315. [http://dx.doi.org/10.1016/0001-6918\(69\)90055-9](http://dx.doi.org/10.1016/0001-6918(69)90055-9).
- Sternberg, S. (2013). The meaning of additive reaction-time effects: Some misconceptions. *Frontiers in Psychology*, 4. <http://dx.doi.org/10.3389/fpsyg.2013.00744>.
- Stoffels, E. J., Van der Molen, M. W., & Keuss, P. J. (1985). Intersensory facilitation and inhibition: Immediate arousal and location effects of auditory noise on visual choice reaction time. *Acta Psychologica*, 58, 45–62.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. <http://dx.doi.org/10.1037/h0054651>.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. London: Cambridge University Press.
- van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavioral Research Methods*, 38, 584–589.
- van Duren, L. L., & Sanders, A. F. (1988). On the robustness of the additive factors stage structure in blocked and mixed choice reaction designs. *Acta Psychologica*, 69, 83–94. doi:0001-6918.
- van Maanen, L., Grasman, R. P., Forstmann, B. U., & Wagenmakers, E. J. (2012). Pieron's law and optimal behavior in perceptual decision-making. *Frontiers in Neuroscience*, 5, 143. <http://dx.doi.org/10.3389/fnins.2011.00143>.
- Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114, 830–841. <http://dx.doi.org/10.1037/0033-295X.114.3.830>.
- Wald, A. (1947). *Sequential analysis*. New York: Wiley.
- Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19, 326–339.
- White, C. N., Brown, S., & Ratcliff, R. (2011). A test of Bayesian observer models of processing in the Eriksen flanker task. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 489–497. <http://dx.doi.org/10.1037/a0026065>.

- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*, 210–238.
- Winer, B. J. (1971). *Statistical principles in experimental design: Design and analysis of factorial experiments*. New York: McGraw-Hill.
- Wylie, S. A., Claassen, D. O., Huizenga, H. M., Schewel, K. D., Ridderinkhof, K. R., Bashore, T. R., et al (2012). Dopamine agonists and the suppression of impulsive motor actions in Parkinson disease. *Journal of Cognitive Neuroscience*, *24*, 1709–1724. http://dx.doi.org/10.1162/jocn_a_00241.
- Wylie, S. A., Ridderinkhof, K. R., Bashore, T. R., & van den Wildenberg, W. P. (2010). The effect of Parkinson's disease on the dynamics of on-line and proactive cognitive control during action selection. *Journal of Cognitive Neuroscience*, *22*, 2058–2073. <http://dx.doi.org/10.1162/jocn.2009.21326>.