Journal of Mathematical Psychology 56 (2012) 208–216

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

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

Notes and comment

How to measure post-error slowing: A confound and a simple solution

Gilles Dutilh [a,*], Don van Ravenzwaaij [a], Sander Nieuwenhuis [b], Han L.J. van der Maas [a], Birte U. Forstmann [a], Eric-Jan Wagenmakers [a]

[a] University of Amsterdam, The Netherlands
[b] Leiden University, The Netherlands

ARTICLE INFO

Article history:
Received 6 August 2011
Received in revised form
27 March 2012
Available online 14 May 2012

Keywords:
Error-monitoring
Post-error slowing
Diffusion model
Sequential effects
Speeded Decision-making

ABSTRACT

In many response time tasks, people slow down after they make an error. This phenomenon of post-error slowing (PES) is thought to reflect an increase in response caution, that is, a heightening of response thresholds in order to increase the probability of a correct response at the expense of response speed. In many empirical studies, PES is quantified as the difference in response time (RT) between post-error trials and post-correct trials. Here we demonstrate that this standard measurement method is prone to contamination by global fluctuations in performance over the course of an experiment. Diffusion model simulations show how global fluctuations in performance can cause either spurious detection of PES or masking of PES. Both confounds are highly undesirable and can be eliminated by a simple solution: quantify PES as the difference in RT between post-error trials and the associated pre-error trials. Experimental data are used as an empirical illustration.

1. Introduction

People tend to slow down after they commit an error, a phenomenon known as post-error slowing (PES). Ever since the classic article "What does a man do after he makes an error?" (Rabbitt, 1966), the PES phenomenon has received considerable attention in the response time (RT) literature and several explanations have been proposed to explain its existence (e.g, Laming, 1968, 1979, Notebaert et al., 2009, Rabbitt & Rodgers, 1977, see Dutilh, Vandekerckhove, Forstmann, & Wagenmakers, 2012), for an empirical comparison). The most popular account of PES states that it reflects an error-induced increase in response caution that allows a participant to maintain a relatively constant level of accuracy (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001, Smith & Brewer, 1995).

Specifically, this account holds that participants continually monitor their performance and interpret errors as a sign that the chosen response threshold was too liberal. Consequently, participants heighten their threshold following an error in order to increase the probability of a correct response on the next trial. The heightened threshold leads to fewer errors but also causes slower responding (i.e., the PES phenomenon).

At the same time, participants interpret correct responses as a sign that the chosen response threshold was too conservative, and therefore they are assumed to lower their threshold following each correct response. Thus, participants become more cautious after an error and slightly more daring after a correct response; in this way the system self-regulates to a state of homeostasis characterized by fast responses and few errors. Fig. 1, based on fictitious but representative data, illustrates the typical pattern of modest post-correct speed-up and pronounced post-error slowing (e.g., Brewer & Smith, 1989; Smith & Brewer, 1995).

This response-monitoring interpretation of PES suggests that the amplitude of PES can be used as a direct measure of cognitive control.[1] Although the response monitoring/cognitive control interpretation might not be appropriate in all cases (e.g., Dutilh, Forstmann, Vandekerckhove, & Wagenmakers, submitted for publication, Notebaert et al., 2009), in many studies it is assumed to be correct from the outset. Consequently, the magnitude of PES is often treated as an important dependent variable that is correlated with neurophysiological variables such as anterior cingulate activity (Danielmeier, Eichele, Forstmann, Tittgemeyer, & Ullsperger, 2011; Li, Huang, Constable, & Sinha, 2006), error-related negativity (ERN) and positivity (Pe; Hajcak, McDonald, & Simons, 2003b), and cortisol levels (e.g., Tops & Boksem, 2010).

* Correspondence to: University of Basel, Department of Psychology, Missionsstrasse 62a, 4055 Basel, Switzerland.
E-mail address: gilles.dutilh@gmail.com (G. Dutilh).

[1] Note that Brewer and Smith (1989), Rabbitt (1979), and Smith and Brewer (1995) interpret the coarseness of the fluctuations of RT around errors as a *negative* indicator of cognitive control. These authors argued that the elderly have coarser control over their speed–accuracy trade-off, indicated by larger fluctuations in RT surrounding an error.

0022-2496/$ – see front matter
doi:10.1016/j.jmp.2012.04.001

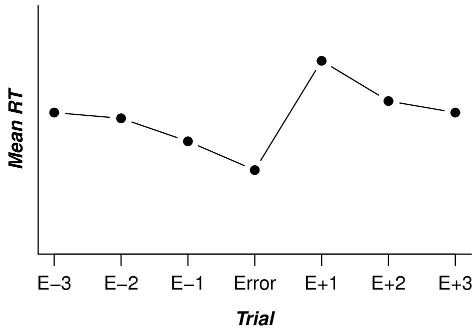

Fig. 1. Typical fluctuations in mean RT surrounding an error E. Fictitious data are representative of results reported in Laming (1979) and Smith and Brewer (1995). Participants tend to speed up until they commit an error, after which they slow down considerably. Subsequent to the post-error trial $E + 1$, participants start to speed up again.

In this article we discuss how PES can best be measured. First we explain how, although straightforward and intuitive, the traditional method to quantify PES can create spurious PES or mask real PES as a result of global changes in performance. We illustrate this confound with two simulation studies and then show how the confound can be eliminated. The final section illustrates both spurious and masked PES in a real data set.

2. The measurement of post-error slowing

There are several methods to quantify PES. The most insightful method plots the fluctuations in mean RT surrounding an error (e.g., Brewer & Smith, 1989; Smith & Brewer, 1995; see Fig. 1 for an example). The resulting graph shows mean RT for error trial E, mean RT for subsequent trials $E + 1$, $E + 2$, etc., and mean RT for preceding trials $E - 1$, $E - 2$, etc. The form of the graph depends slightly on what trials are included in the calculations. For example, one may choose to include pre-error trials that are also post-error trials, one may include errors that are simultaneously pre-error or post-error trials, and so forth. Figs. 4–6 are based on seven-trial sequences where the middle trial was the sole error, ensuring that fluctuations in mean RT are not confounded by fluctuations in error rate. Such a selection of trial sequences may not be feasible when the number of observations is low.

Although the graphical method is very informative, researchers often prefer a method that quantifies the magnitude of PES in a single number. The traditional and most intuitive method to quantify PES in a single number is to calculate the difference in mean RT (MRT) between trials post-error and trials post-correct. This difference, $\widehat{\text{PES}}_{\text{traditional}} = \text{MRT}_{\text{post-error}} - \text{MRT}_{\text{post-correct}}$, is often calculated per condition per participant and is used as a behavioral variable for further analysis. The magnitude of $\widehat{\text{PES}}_{\text{traditional}}$ may depend slightly on whether or not error trials are included in the calculation of $\text{MRT}_{\text{post-error}}$ and $\text{MRT}_{\text{post-correct}}$ (e.g., Hajcak & Simons, 2008).[2]

The popularity of the traditional method is highlighted in Table 1, which lists all 14 articles that quantify PES and have over 100 citations.[3] The right column of Table 1 indicates the method used to quantify PES. Nine out of the 14 articles used the traditional measure described above. Three articles used an adjusted version of the traditional method that is sensitive to the same problem that we outline below.

3. A confound

The traditional method of quantifying PES, $\widehat{\text{PES}}_{\text{traditional}} = \text{MRT}_{\text{post-error}} - \text{MRT}_{\text{post-correct}}$, has strong face validity. However, the method is vulnerable to a confound that was already hinted at by Laming (1979, p. 205) when he suggested …

> …the possibility that errors and the increased RT on trials which follow them are jointly due to a local deterioration in performance. Suppose, for example, that the subject suffers short periods of relative inattention to the CR [choice response] task …During these periods RTs are longer and errors more frequent than normal.

Such local deterioration of performance leads to two possible complications when PES is calculated using the traditional method. The first complication is that global changes in *ability* or *motivation* may lead to spurious post-error *slowing*. The second complication is that global changes in *response caution* may lead to spurious post-error *speed-up*. Both situations are illustrated in Fig. 2.

First, consider the hypothetical scenario where a participant starts a one-hour, 1000-trial experimental session with high motivation. As a result, the participant's responses are fast and accurate. However, as the session proceeds, fatigue starts to kick in and motivation drops. This decrease in motivation is illustrated in the upper left panel of Fig. 2. With low motivation, the participant's responses become increasingly slow (middle left panel) and inaccurate (bottom left panel). Now suppose that this participant does not slow down after errors, that is, real PES is completely absent. Now, we quantify PES in this participant's data with the standard method $\widehat{\text{PES}}_{\text{traditional}} = \text{MRT}_{\text{post-error}} - \text{MRT}_{\text{post-correct}}$. Notice that most post-error RTs will originate from the second half of the session, because there are more errors in that half. Likewise, most post-correct RTs will originate from the first half of the session, because there are more correct responses in that half. Because of the decrease in motivation, responses in the first half of the experiment are quicker than responses in the second half of the experiment. Therefore, post-correct trials will on average be faster than post-error trials, despite the fact that there is no real PES. Thus, the traditional comparison of post-error RTs with post-correct RTs can yield an artificial PES effect (or even mask post-error speed-up).[4]

Second, consider the hypothetical scenario where a participant starts an experimental session very keen on being accurate. The participant's responses are then highly accurate but slow. As the session continues, the participant may get increasingly careless. The associated decrease in response caution is illustrated in the upper right panel of Fig. 2. With decreasing caution, the participant's responses become quicker (middle right panel) at the cost of accuracy (bottom right panel). Analogous to the previous case, suppose that for this participant, real PES is completely absent. Once again we analyze PES in this participant's data set in the standard fashion, that is, $\widehat{\text{PES}}_{\text{traditional}} = \text{MRT}_{\text{post-error}} - \text{MRT}_{\text{post-correct}}$. Note that most error RTs and therefore most post-error RTs will again originate from the second half of the experiment, where the participant was rather careless. As before, most post-correct RTs originate from the first half of the experiment, where the participant was relatively careful. In contrast to the first scenario, however, responses are faster in the second half of the experiment than in the first half. Therefore, post-correct trials are on average slower than post-error trials, despite the fact that there is no real PES. Thus, the traditional comparison

[2] A related method compares (correct) post-error trials to all correct trials (both post-error and post-correct). This method is also vulnerable to the confound of global fluctuations in performance that we discuss in this paper.

[3] Obtained from scholar.google.com, December 2011.

[4] A similar artificial PES effect occurs when participants gradually improve on the task at hand, for instance through practice.

Table 1
Popularity of the traditional measure for PES. Included are all 14 articles that quantify PES and have more than 100 citations. This list is based on a scholar.google.com search on "post-error slowing". Nine out of 14 articles use the traditional method, and only one article reported differences in mean RT between pre-error vs. error trials and between post-error vs. error trials. Three studies (adj.) matched correct trials to error trials before comparing the post-correct to post-error trials in order to control for the fact that errors are often faster than correct responses. This adjustment does not correct for the confound described in this article.

Reference	Number of citations	Measure
Hajcak et al. (2003b)	167	$\widehat{\text{PES}}_{\text{traditional}}$
Kerns et al. (2004)	1014	$\widehat{\text{PES}}_{\text{traditional}}$
Gehring and Knights (2000)	563	$\widehat{\text{PES}}_{\text{traditional}}$
Klein et al. (2007)	112	$\widehat{\text{PES}}_{\text{traditional}}$
Hajcak et al. (2004)	141	$\widehat{\text{PES}}_{\text{traditional}}$ (adj.)
Hajcak et al. (2003a)	159	$\widehat{\text{PES}}_{\text{traditional}}$ (adj.)
Hajcak and Simons (2002)	158	$\widehat{\text{PES}}_{\text{traditional}}$ (adj.)
Alain et al. (2002)	106	$\widehat{\text{PES}}_{\text{traditional}}$
De Bruijn et al. (2004)	117	$\widehat{\text{PES}}_{\text{traditional}}$
Fellows and Farah (2005)	118	$\widehat{\text{PES}}_{\text{traditional}}$
Egner and Hirsch (2005)	148	$\text{MRT}_{\text{post-error}} - \text{MRT}_{\text{correct}}$
Jones et al. (2003)	117	$\widehat{\text{PES}}_{\text{traditional}}$
Mathalon et al. (2002)	141	$\widehat{\text{PES}}_{\text{traditional}}$
Stuss et al. (2003)	159	$\text{MRT}_{\text{post-error}} - \text{MRT}_{\text{error}}$, $\text{MRT}_{\text{pre-error}} - \text{MRT}_{\text{error}}$

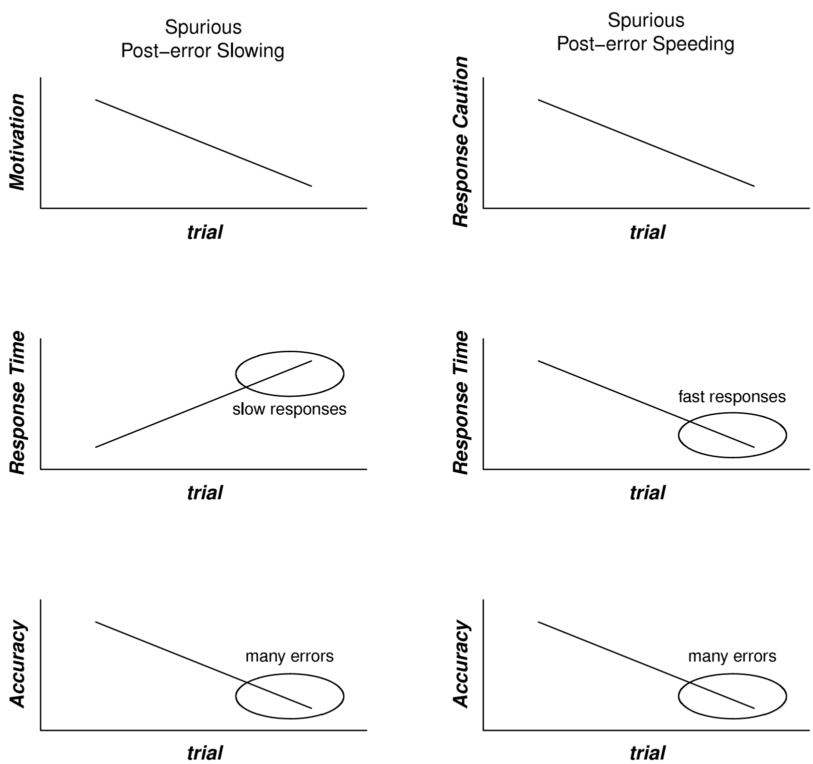

Fig. 2. Schematic representation of two confounds. The three panels in the left column illustrate that, when a participant's motivation decreases during an experimental session, slow RTs co-occur with *low* accuracy. In this case, calculation of $\widehat{\text{PES}}_{\text{traditional}}$ may result in spurious or inflated estimates of PES. The three panels in the right column illustrate that, when a participant's caution decreases during an experimental session, slow RTs co-occur with *high* accuracy. In this case, calculation of $\widehat{\text{PES}}_{\text{traditional}}$ may result in spurious post-error speeded, or deflated estimates of PES.

of post-error RTs with post-correct RTs can yield an artificial post-error speed-up effect (or mask PES when it is present).

The two scenarios described above show that global changes in performance may systematically confound the estimation of PES, at least when it is quantified as $\widehat{\text{PES}}_{\text{traditional}}$. Below we present two simulation studies that support this conclusion by quantifying the impact of the two confounds using a process model of RT.

4. Simulation studies

In two simulation studies we used the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008; Wagenmakers, 2009) to make the scenarios described above more concrete. The diffusion model produces both RTs and percentage correct. Most importantly for this study, the model can describe the specific influences of motivation and response caution on response time data.

4.1. The diffusion model

In the diffusion model for speeded two-choice tasks (Ratcliff, 1978), stimulus processing is modeled as the noisy accumulation of evidence over time. A response is initiated when the accumulated evidence reaches a predefined boundary (Fig. 3). The main components of the diffusion model are (1) speed of information processing, quantified by drift rate v. Low absolute values of v produce relatively long RTs and high error rates; (2) response

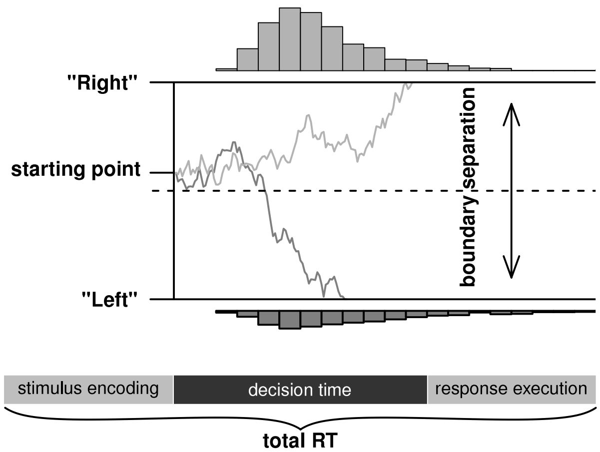

Fig. 3. Graphical illustration of the diffusion model, as applied to the random dot motion task. In this task, participants have to decide quickly whether a stimulus appears to move left or right. Consider a trial in which the stimulus moves to the right. The two example sample paths represent the accumulation of evidence which result in one correct response ("right", light line) and one error response ("left", dark line). Repeated application of the diffusion process yields histograms of both correct responses (upper histogram) and incorrect responses (lower histogram). As is evident from the histograms, the correct, upper "right" boundary is reached more often than the incorrect, lower "left" boundary. The total RT consists of the sum of a decision component, modeled by the noisy accumulation of evidence, and a nondecision component that represents the time needed for processes such as stimulus encoding and response execution.

caution, quantified by boundary separation a. Low values of a lead to relatively short RTs and high error rates; and (3) *a priori* bias, quantified by starting point z. Together, these parameters generate a distribution of decision times DT. The observed RT, however, also consists of stimulus-nonspecific components such as response preparation and motor execution, which together make up nondecision time T_{er}. The model assumes that T_{er} simply shifts the distribution of DT, such that $RT = \mathrm{DT} + T_{er}$ (Luce, 1986). The model specification is completed by including parameters that specify across-trial variabilities in drift rate, starting point, and nondecision time (Ratcliff & Tuerlinckx, 2002). These variability parameters allow the model to account for empirical phenomena such as the finding that errors can be systematically faster or systematically slower than correct responses.

After the diffusion model has been fit to data, the parameters can be interpreted in terms of psychological processes (e.g., ability, caution, bias), allowing a decomposition of performance into its constituent cognitive elements. The diffusion model has been applied to a range of different tasks such as lexical decision (e.g., Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009) and perceptual discrimination (e.g., Ratcliff & Rouder, 1998) with the goal to increase our knowledge of aging (e.g., Ratcliff, Thapar, & McKoon, 2006, 2010), practice (e.g., Dutilh, Krypotos, & Wagenmakers, 2011; Dutilh et al., 2009), and clinical disorders (e.g., White, Ratcliff, Vasey, & McKoon, 2009, 2010a,b).

In the simulation studies, we use the diffusion model to simulate plausible RT data that reflect the two hypothetical scenarios described in the previous section. For the first scenario, in which motivation changes systematically across trials, we simulated data in which all parameters of the diffusion model were constant, but drift rate v changed. For the second scenario, in which response caution changes systematically across trials, we simulated data in which all parameters of the diffusion model were constant, but boundary separation a changed.

4.2. Simulation I: spurious post-error slowing

In this section we reproduce the first scenario described above: spurious PES as a consequence of global changes in motivation. In order to minimize the effects of sampling variation, we simulated an experiment of 100,000 consecutive trials.[5] Data were generated from the diffusion model using a default set of plausible parameters (Matzke & Wagenmakers, 2009): $a = 0.12, z = a/2, T_{er} = 0.300, s_z = 0.2a$ (i.e., across-trial variability in starting point), $\eta = 0$ (i.e., across-trial variability in drift rate), and $s_t = 0$ (i.e., across-trial variability in nondecision time). We used drift rate as a proxy for motivation. Note that in this simulation the PES effect was completely absent.

We explored the effect of four types of fluctuation in motivation (i.e., drift rate): drift rate constant, drift rate decreasing as a step function, drift rate decreasing linearly, and drift rate fluctuating as a sine function. The simulation results are shown in Fig. 4 with one panel for each type of fluctuation (as denoted in figure headers). Each panel shows mean error RT, and mean correct RT for the three trials preceding an error and the three trials following an error. The asterisk represents mean correct RT for post-correct trials. The vertical gray bars quantify the traditional measure for PES, that is, $\widehat{\mathrm{PES}}_{\mathrm{traditional}} = \mathrm{MRT}_{\text{post-error}} - \mathrm{MRT}_{\text{post-correct}}$.

In the top-left panel of Fig. 4 drift rate is a constant $v = 0.22$ across trials. Consequently, mean RT is also constant and does not depend on the position relative to the error E. The measure $\widehat{\mathrm{PES}}_{\mathrm{traditional}}$ appropriately indicates a negligible PES effect.

Now consider the upper right panel of Fig. 4. The only difference with the upper left panel is that drift rate v is now linearly decreasing from $v_{\mathrm{high}} = 0.38$ on the first trial to $v_{\mathrm{low}} = 0.06$ on the last trial. Again, mean RT around errors is practically constant, suggesting (correctly) that there is no real PES. However, the traditional measure $\widehat{\mathrm{PES}}_{\mathrm{traditional}}$ now yields a spurious PES effect of 53 ms. The bottom left panel shows the results of a similar simulation where drift rate is high and constant in the first half of the experiment (i.e., $v_{\mathrm{high}} = 0.38$) and low and constant in the second half of the experiment (i.e., $v_{\mathrm{low}} = 0.06$). The bottom right panel displays the results of a simulation where drift rate was fluctuating according to a sine wave (with a period of 100 trials; $\max(v) = v_{\mathrm{high}} = 0.38$; $\min(v) = v_{\mathrm{low}} = 0.06$). Just as in the simulation with a linear change in drift rate, mean RT around errors is practically constant, suggesting (correctly) that there is no real PES. However, the traditional measure $\widehat{\mathrm{PES}}_{\mathrm{traditional}}$ yields spurious PES effects. Fig. 4 also shows the results of the robust measure for PES that we discuss later. This method appropriately indicates that there is no PES.

In sum, these simulations demonstrate that a systematic change in motivation confounds $\widehat{\mathrm{PES}}_{\mathrm{traditional}}$, possibly leading to spurious PES. Note that the direction of change is irrelevant, since both an increase and a decrease in motivation yield similar local differences.

4.3. Simulation II: masked post-error slowing

In this section we reproduce the second scenario described earlier: masked PES as a consequence of global changes in caution. In order to minimize the effects of sampling variation, we again simulated an experiment of 100,000 consecutive trials. Data were generated from the diffusion model using $v = 0.22, z = a/2, T_{er} = 0.300, s_z = 0.2a$ (i.e., across-trial variability in starting point), $\eta = 0$ (i.e., across-trial variability in drift rate), and $s_t = 0$ (i.e., across-trial variability in nondecision time). We used boundary separation as a proxy for caution. In this simulation, we generated a PES effect by increasing boundary separation after an error. Specifically, all trials following a correct response used a lower boundary separation of $a(t) - 0.008$ (where $a(t)$ indicates the local level of boundary separation that may be subject to

[5] For each panel shown in Fig. 4 we simulated one experiment.

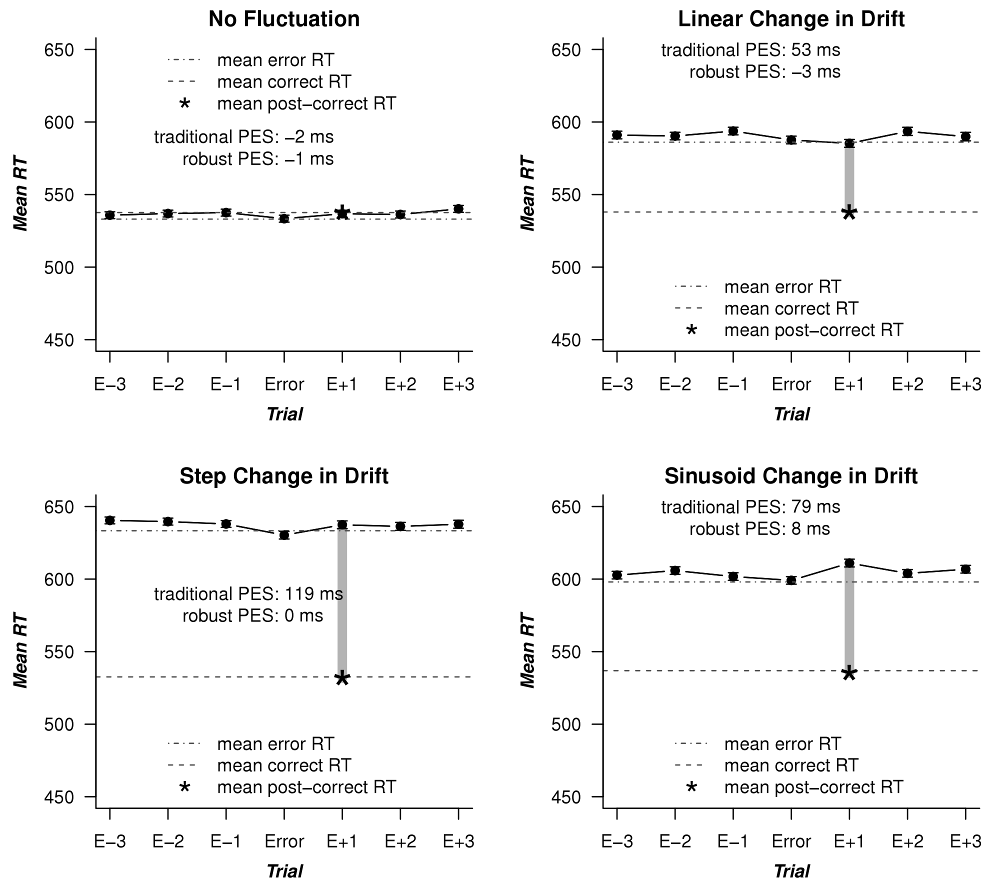

Fig. 4. Performance of two PES measures when drift rate fluctuates over trials and PES is absent. The solid line represents mean RT for correct trials around an error and for the error itself. When drift rate fluctuates systematically across trials (bottom-left, as a step function; top-right, linearly; bottom-right, as a sine function), the traditional measure leads to spurious detection of PES (the vertical gray bars). In contrast, the robust measure supports the correct conclusion: there is no PES. Error bars indicate one standard error of the mean.

systematic variation as indicated below), and all trials following an error used a higher boundary separation of $a(t) + 0.008$.

We explored the effect of four types of fluctuation in caution (i.e., boundary separation): boundary separation constant, boundary separation decreasing as a step function, boundary separation decreasing linearly, and boundary separation fluctuating as a sine function. The simulation results are shown in Fig. 5 with one panel for each type of fluctuation (as denoted in figure headers). As before, each panel shows mean error RT, and mean correct RT for the three trials preceding an error and the three trials following an error. The asterisk represents mean correct RT for post-correct trials. The vertical gray bars again quantify the traditional measure for PES, that is, $\widehat{\text{PES}}_{\text{traditional}} = \text{MRT}_{\text{post-error}} - \text{MRT}_{\text{post-correct}}$.

In the top-left panel of Fig. 5, boundary separation is a constant $a(t) = 0.18$ across trials. Mean RT depends on the position relative to the error E, and the measure $\widehat{\text{PES}}_{\text{traditional}}$ appropriately indicates a PES effect of 34 ms.

Now consider the upper right panel of Fig. 5. The only difference with the upper left panel is that boundary separation $a(t)$ is now linearly decreasing from $a_{\text{high}} = 0.22$ on the first trial to $a_{\text{low}} = 0.14$ on the last trial. Again, mean RT around errors is not constant, suggesting (correctly) that there is PES. However, the traditional measure $\widehat{\text{PES}}_{\text{traditional}}$ reports a deflated PES effect of only 6 ms. The bottom left panel shows the results of a similar simulation where boundary separation is high and constant in the first half of the experiment (i.e., $a_{\text{high}} = 0.22$) and low and constant in the second half of the experiment (i.e., $a_{\text{low}} = 0.14$). The bottom right panel displays the results of a simulation where boundary separation was fluctuating according to a sine wave (with a period of 100 trials; $\max(a) = a_{\text{high}} = 0.22$; $\min(a) = a_{\text{low}} = 0.14$). Just as in the simulation with a linear change in boundary separation, mean RT around errors is not constant, suggesting (correctly) that there is PES. However, the traditional measure $\widehat{\text{PES}}_{\text{traditional}}$ yields very small or even negative PES effects (i.e., post-error speed-up). Fig. 5 also shows that the robust measure for PES that we discuss later appropriately indicates that there is indeed PES.

In sum, these simulations demonstrate that a systematic change in caution confounds $\widehat{\text{PES}}_{\text{traditional}}$, possibly leading to an underestimated or masked PES. Note that the direction of change is again irrelevant, since both an increase and a decrease in response caution yield similar local differences.

5. A simple solution

The above confounds arise because post-correct and post-error trials (i.e., the trials used to calculate $\widehat{\text{PES}}_{\text{traditional}}$) are not evenly distributed across the time series. The confounds can be eliminated when we compare post-error trials to post-correct trials that originate from the same locations in the time series. One natural option is to use post-correct trials that are *pre-error* trials at the same time. So, instead of comparing the mean RTs of all post-error trials to those of all post-correct trials, we conduct pairwise comparisons around each error (see Nelson, Boucher, Logan, Palmeri, & Schall, 2010 for a similar method in a different context). In other words, we take the average of $RT(E + 1) - RT(E-1)$ for all errors E. This comparison will now be referred to as the *robust* measure $\widehat{\text{PES}}_{\text{robust}}$ as opposed to the *traditional* measure $\widehat{\text{PES}}_{\text{traditional}}$.

Figs. 4 and 5 allow a comparison of $\widehat{\text{PES}}_{\text{robust}}$ and $\widehat{\text{PES}}_{\text{traditional}}$. As anticipated, $\widehat{\text{PES}}_{\text{robust}}$ is not much affected by global fluctuation

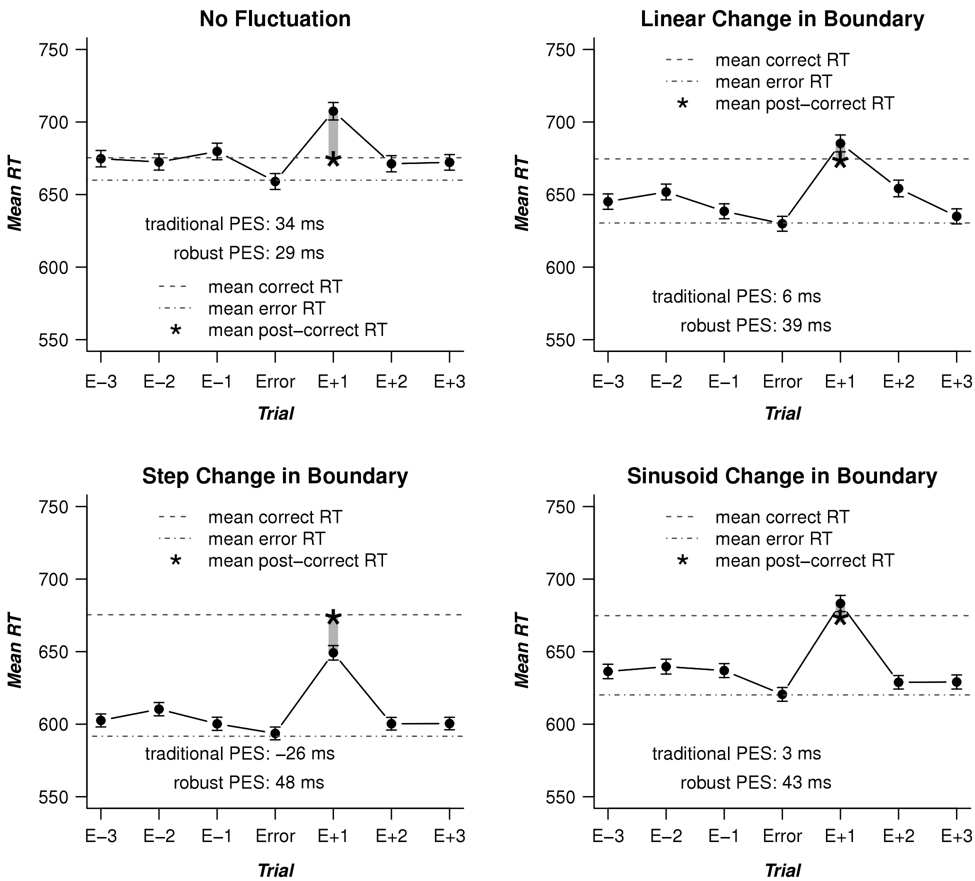

Fig. 5. Performance of two PES measures when boundary separation fluctuates over trials and PES is present. The solid line represents mean RT for correct trials around an error and for the error itself. When boundary separation fluctuates systematically across trials (bottom-left, as a step function; top-right, linearly; bottom-right, as a sine function), the traditional measure fails to detect positive PES (the vertical gray bars). In contrast, the robust measure supports the correct conclusion: there is PES. Error bars indicate one standard error of the mean.

in motivation and caution, but $\widehat{\text{PES}}_{\text{traditional}}$ is. For the simulations without PES but with relatively pronounced fluctuations in motivation, $\widehat{\text{PES}}_{\text{robust}}$ correctly estimates PES to be very small; for the simulations with PES but with relatively pronounced fluctuations in caution, $\widehat{\text{PES}}_{\text{robust}}$ correctly estimates PES to be in between 29 and 48 ms.[6] This behavior is the mirror opposite of $\widehat{\text{PES}}_{\text{traditional}}$, a measure that consistently arrives at the incorrect conclusion.

6. Towards a principled solution

The measure $\widehat{\text{PES}}_{\text{robust}}$ is designed to be immune against global performance fluctuations that may adversely affect the widely used measure $\widehat{\text{PES}}_{\text{traditional}}$. However, two important problems remain. First, PES measures capture error-induced changes in a specific variable, namely mean RT; the measures ignore changes in accuracy and changes in RT distributions. Such changes can be captured by the diffusion model described above. The diffusion model, however, currently does not have a mechanism for trial-to-trial adjustments in the relevant psychological processes such as response caution. Second, PES measures are difficult to interpret in terms of cognitive control—do participants with pronounced PES differ from other participants because they monitor their performance more closely, or because they exert coarser control over their behavior?

In order to quantify the magnitude of cognitive control, it should first be established how this magnitude relates to the observed variables RT and accuracy. One attempt to formalize this relation was made by Vickers (1979), who proposed a self-regulating accumulator model of decision making. The model describes how participants keep track of the confidence in each response. This confidence level is than contrasted to a target level of confidence. Based on this contrast, participants adjust their level of response caution after each trial. Importantly, the model has two specific parameters that govern two important aspects of cognitive control: one parameter reflects how carefully participants monitor performance; a second parameter describes how coarsely they adjust their behavior.[7] Another attempt to formally describe how participants adjust their behavior to meet different task constraints is made by Simen, Cohen, and Holmes (2006), who propose a neurologically plausible model of how participants' estimates of reward rates are translated to a dynamic setting of response thresholds.

The foregoing illustrates how a comprehensive measure of cognitive control should be based on a formal model that describes how speed and accuracy on a given trial depend on performance for earlier trials. Therefore, a comprehensive measure should explicitly describe intact sequences of trials, rather than the properties of different independent trial conditions such as post-error and post-correct. Note that the traditional method to quantify

6 Recall that the PES effect was simulated as an increase in boundary separation. The resulting effect on mean RT is moderated by the fluctuations in baseline boundary separation across trials.

7 Note that the self-regulating accumulator model does not attribute any special role to errors. Nonetheless, PES is naturally accounted for by the hypothesized process.

PES depends on the implicit assumption that trials are independent instances of different conditions. In contrast, the method that we propose treats the trials around errors as parts of intact time series. Therefore, relative to the traditional measure of PES, the robust measure we propose here more closely approximates a principled measure of cognitive control.

7. Empirical Illustration: the confound is real

The simulations above showed that $\widehat{\text{PES}}_{\text{traditional}}$ can detect spurious PES and mask real PES. We now provide an empirical illustration of these two situations. For this illustration, we selected data from two individual participants from a larger study that will be published elsewhere.

7.1. Method

Elderly participants were presented with the random dot motion task Britten, Shadlen, Newsome, and Movshon (1992), a task that is popular in cognitive neuroscience and research on monkeys. The random dot stimulus consists of a circular display of dots. The dots appear, disappear and are replaced in such a way that the entire circle of dots appears to move either left or right. The apparent motion that the participant perceives can be best described as the flickering of a turning disco ball in a spotlight. This illusion is created as follows. At each frame (50 ms), 120 dots are displayed. Every next frame, an experimentally defined proportion P_{move} of the dots from the former frame are shifted a certain distance l_{move} to the target side (e.g., right, if the correct response is right). The remaining portion of the pixels is randomly replaced in the circle (independent of their previous positions). P_{move} was set to 50%. l_{move} was set to one pixel. Participants were instructed to indicate the direction of the apparent movement (left or right) by pressing one of two response buttons. For the data presented here, participants were asked to respond accurately, but not waste too much time. Two participants were selected for the current empirical illustration.

7.2. Results

Data for the two participants are shown in Fig. 6. The interpretation of these panels follows that of Figs. 4 and 5; each panel shows the mean RT of correct trials that either precede ($E-1, E-2, E-3$) or follow ($E+1, E+2, E+3$) an error. The asterisk again represents the mean post-correct RT, and the vertical gray bars indicates the traditional measure of PES, that is, $\widehat{\text{PES}}_{\text{traditional}} = \text{MRT}_{\text{post-error}} - \text{MRT}_{\text{post-correct}}$.

Participant A, whose data ($n = 1400$ trials) are displayed in the upper panel of Fig. 6, shows a pattern similar to the synthetic patterns shown in Fig. 4. The traditional measure estimates a sizable 90 ms. PES effect but the robust measure suggests PES is entirely absent. The latter conclusion is supported by the relatively constant mean RT around errors. This data set is an example of spurious PES.

Participant B, whose data ($n = 1300$ trials) are displayed in the lower panel of Fig. 6, shows a pattern similar to the synthetic patterns shown in Fig. 5.[8] The traditional measure estimates a modest 25 ms PES effect but the robust measure suggests that the true PES is much larger (i.e., 88 ms.). The latter conclusion is supported by the fact that the mean RT changes sharply around errors. This data set is an example of masked PES.

[8] The two participants have unequal number of trials because the testing protocol was to collect as many data as possible in a fixed time period.

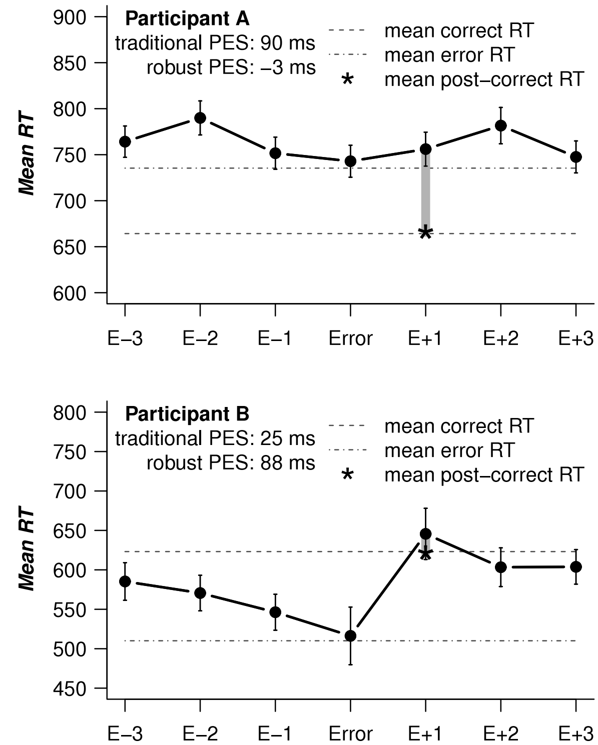

Fig. 6. Fluctuation of mean correct RT around error trials (line) and post-correct RT (asterisk). Each panel shows data for one selected participant. The upper panel participant shows no PES, but the traditional measure falsely indicates the presence of PES. The lower panel participant does show PES, but the traditional measure strongly underestimates the effect. Error bars indicate one standard error of the mean.

Fig. 7 shows, for each of the two participants discussed above, the entire time series as a 11-trial moving average of RT. Error responses are indicated as black dots.

The upper panel shows that participant A has large fluctuations in both RT and error rate; moreover, the time series shows phases of poor performance (i.e., long RTs and high error rates). The lower panel shows that participant B has large fluctuations in RT. For both participants, the pronounced temporal fluctuations are the most likely explanation for the difference between the robust and traditional measures of PES. It should be mentioned that the response time distributions of both participants look regular and that the percentages correct (81% and 99%, respectively) are not unusual in the moving dots task.

8. Concluding comments

Over the last two decades, cognitive control has become a major research topic in experimental psychology. PES is assumed to be an indicator of cognitive control and as such it is often used as an important behavioral variable. Recently, however, some studies have questioned whether PES really reflects cognitive control (e.g., Dutilh et al., 2012; Notebaert et al., 2009). In this study, however, we focused on a more elementary issue regarding the application of PES as a dependent variable: the reliability of its estimation.

The traditional method to quantify PES compares mean RT after errors with mean RT after correct responses. We demonstrated that this analysis is vulnerable to two confounds that both result from fluctuations of a participant's performance over the course of an experiment. These confounds are global changes in motivation or ability and global changes in response caution. We showed how these confounds may lead to both spurious and masked PES effects, both in simulations and in two empirical data sets. The solution we offered is both simple and adequate: in the comparison of

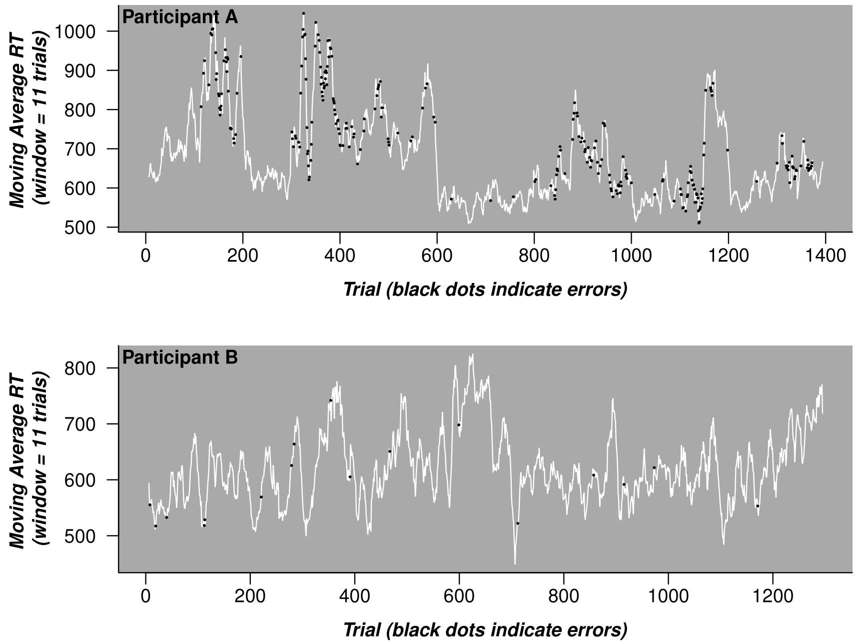

Fig. 7. Response time and error rate fluctuate over trials. The lines show an 11-trial moving average of RT, the black dots indicate where errors are committed. Participant A, for whom we found spurious post-error slowing, shows large fluctuations in both RT and error rate. Participant B, for whom we found masked post-error slowing, shows large fluctuations in RT.

post-error mean RT and post-correct mean RT, only those post-correct trials should be included that are also pre-error trials. This additional condition of our *robust* method ensures that post-error and post-correct trials originate from the same locations in the data set. This property of the robust method makes it immune to global fluctuations in performance.

The field of research that uses PES as a behavioral variable has expanded strongly over the past decade. The PES phenomenon itself however has not gained interest proportionally. Consequently, little attention has been given to the theoretical background of PES, let alone the related methodological issues. There seems to be no compelling reason why researchers have stuck to the traditional method of quantification. The robust method we propose here is easy to apply, insensitive to changes in performance over the course of an experiment, and hence allows a more informative and veridical measurement of post-error slowing. Our robust method does however not describe *why* participants slow down after an error. Furthermore, it focuses on mean RT and ignores accuracy and the spread of RT. A comprehensive measure of cognitive control should follow from a model that describes how the interplay of a decision maker's ability and cognitive control produces the time series of accuracy and RT.

References

Alain, C., McNeely, H. E., He, Y., Christensen, B. K., & West, R. (2002). Neurophysiological evidence of error-monitoring deficits in patients with schizophrenia. *Cerebral Cortex*, *12*, 840–846.

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

Brewer, N., & Smith, G. A. (1989). Developmental changes in processing speed: influence of speed-accuracy regulation. *Journal of Experimental Psychology*, *118*, 298–310.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*, 4745–4765.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *The Journal of Neuroscience*, *31*, 1780.

De Bruijn, E. R. A., Hulstijn, W., Verkes, R. J., Ruigt, G. S. F., & Sabbe, B. G. C. (2004). Drug-induced stimulation and suppression of action monitoring in healthy volunteers. *Psychopharmacology*, *177*, 151–160.

Dutilh, G., Forstmann, B.U., Vandekerckhove, J., & Wagenmakers, E.J. (2012). A diffusion model account of age differences in post-error slowing. Manuscript (submitted for publication).

Dutilh, G., Krypotos, A. M., & Wagenmakers, E. J. (2011). Task-related vs. stimulus-specific practice: a diffusion model account. *Journal of Experimental Psychology*, *58*, 434–442.

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., & Wagenmakers, E. J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*, 454–465.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E. J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*, 1026–1036.

Egner, T., & Hirsch, J. (2005). The neural correlates and functional integration of cognitive control in a Stroop task. *NeuroImage*, *24*, 539–547.

Fellows, L. K., & Farah, M. J. (2005). Is anterior cingulate cortex necessary for cognitive control? *Brain*, *128*, 788–796.

Gehring, W. J., & Knights, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, *3*, 516–520.

Hajcak, G., McDonald, N., & Simons, R. (2003a). Anxiety and error-related brain activity. *Biological Psychology*, *64*, 77–90.

Hajcak, G., McDonald, N., & Simons, R. F. (2003b). To err is autonomic: error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, *40*, 895–903.

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*, 189–197.

Hajcak, G., & Simons, R. F. (2002). Error-related brain activity in obsessive–compulsive undergraduates. *Psychiatry Research*, *110*, 63–72.

Hajcak, G., & Simons, R. F. (2008). Oops!.. I did it again: an ERP and behavioral study of double-errors. *Brain and Cognition*, *68*, 15–21.

Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science*, *6*, 498–504.

Kerns, J. G., Cohen, J. D., MacDonald, A., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*(5660), 1023–1026.

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., Von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *NeuroImage*, *34*, 1774–1781.

Laming, D. (1968). *Information theory of choice—reaction times*. Academic Press.

Laming, D. (1979). Choice reaction preformance following an error. *Acta Psychologica*, *43*, 199–224.

Li, R. C., Huang, C., Constable, R. T., & Sinha, R. (2006). Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. *Journal of Neuroscience*, *26*, 186–192.

Luce, R. D. (1986). *Response times* (8th ed.). New York, NY: Oxford University Press.

Mathalon, D. H., Fedor, M., Faustman, W. O., Gray, M., Askari, N., & Ford, J. M. (2002). Response-monitoring dysfunction in schizophrenia: an event-related brain potential study. *Journal of Abnormal Psychology*, *111*, 22–41.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychonomic Bulletin & Review*, *16*, 798–817.

Nelson, M. J., Boucher, L., Logan, G. D., Palmeri, T. J., & Schall, J. D. (2010). Nonindependent and nonstationary response times in stopping and stepping saccade tasks. *Attention, Perception, & Psychophysics*, *72*, 1913–1929.

Notebaert, W., Houtman, F., Opstal, F., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition*, *111*, 275–279.
Rabbitt, P. (1966). How old and young subjects monitor and control responses for accuracy and speed. *Journal of Experimental Psychology*, *71*, 264–272.
Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology*, *70*, 305–311.
Rabbitt, P., & Rodgers, B. (1977). What does a man do after an error? an analysis of response programming. *Quarterly Journal of Experimental Psychology*, *29*, 727–743.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.
Ratcliff, R., & Rouder, J. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.
Ratcliff, R., Thapar, J., & McKoon, G. (2006). Aging, practice, and perceptual tasks: a diffusion model analysis. *Psychology and Aging*, *21*, 353–371.
Ratcliff, R., Thapar, J., & McKoon, G. (2010). Individual differences, ageing, and IQ in two-choice tasks. *Cognitive Psychology*, *60*, 127–157.
Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychological Bulletin & Review*, *9*, 438–481.
Simen, P., Cohen, J. D., & Holmes, P. (2006). Rapid decision threshold modulation by reward rate in a neural network. *Neural Networks*, *19*, 1013–1026.
Smith, G. A., & Brewer, N. (1995). Slowness and age: speed-accuracy mechanisms. *Psychology and Aging*, *10*, 238–247.
Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: the frontal lobes control individual performance variability. *Brain*, *126*, 2363–2380.
Tops, M., & Boksem, M. A. (2010). Cortisol involvement in mechanisms of behavioral inhibition. *Psychophysiology*, 1–10.
Vickers, D. (1979). *Decision processes in visual perception*. London: Academic Press.
Wagenmakers, E. J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, *21*, 641–671.
White, C. N., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: a diffusion-model analysis. *Cognition & Emotion*, *23*, 181–205.
White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010a). Anxiety enhances threat processing without competition among multiple inputs: a diffusion model analysis. *Emotion*, *10*, 662–677.
White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010b). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, *54*, 39–52.