Inhibiting Responses to Difficult Choices

Dora Matzke

Department of Psychology, University of Amsterdam, The Netherlands

Samuel Curley

School of Psychology, University of Newcastle, Australia

Charlene Q. Gong

King's College Hospital, London

Andrew Heathcote

School of Medicine, University of Tasmania, Australia

Word count: $\sim 9,800$

Correspondence concerning this article should be addressed to:

Dora Matzke, Department of Psychology, PO Box 15906, 1001 NK Amsterdam, The Netherlands, E-mail: d.matzke@uva.nl.

Abstract

The stop-signal paradigm is a widely used procedure to study response inhibition. It consists of a two-choice response time task (go task) that is occasionally interrupted by a stop signal instructing participants to withhold their response. The paradigm owes its popularity to the underlying race model that enables estimation of the otherwise unobservable latency of stopping. As the race model assumes a single go runner that produces the response unless it is beaten by an inhibitory stop runner, it cannot account for errors on the go task. We propose a parametric framework that extends the standard two-runner race model to account for go errors, and hence expand the scope of the stop-signal paradigm to the study of response inhibition in the context of difficult choices. We combine our treatment of go errors with the ability to address two common contaminants in stop-signal data: failures to trigger the go or the stop runner. We show with simulations that applying two-runner parametric race models to difficult choices can severely bias conclusions about response inhibition. Importantly, we also show that even infrequent errors, which have been common in previous stop-signal studies, can result in underestimation of stopping latencies. We demonstrate that our framework enables researchers to study difficult-choice inhibition even in relatively small samples by applying it to novel stop-signal data with high error rates and a manipulation of task difficulty, showing that it provides an accurate characterization of behavior and precise stop estimates.

Keywords: choice errors, ex-Gaussian distribution, go failures, race model, stop-signal paradigm, trigger failures

Introduction

As a central component of executive control, response inhibition receives considerable attention in experimental psychology (Aron, Robbins, & Poldrack, 2014; Logan, 1994; Miyake et al., 2000; Ridderinkhof, Van Den Wildenberg, Segalowitz, & Carter, 2004). The concept refers to the ability to stop ongoing responses that are no longer appropriate, such as stopping in the middle of an expletive (or preferably earlier) during a job interview. As such, response inhibition facilitates adaptive and goal-directed behavior in dynamic environments. In laboratory settings, response inhibition is most commonly investigated using the stop-signal paradigm (Logan & Cowan, 1984; for reviews, see Logan, 1994; Matzke, Verbruggen, & Logan, in press; Verbruggen & Logan, 2009).

The stop-signal paradigm typically consists of an easy two-choice response time (RT) task, such as responding to the direction of an arrow (e.g., press left button for a left-pointing arrow and right button for a right-pointing arrow). Occasionally, this primary "go" task is interrupted by a stop signal presented on a variable delay (i.e., stop-signal delay [SSD]) that instructs participants to withhold their response on that trial. Response inhibition is successful when the stop signal is presented sufficiently close to the onset of the go stimulus, but it fails when the stop signal is presented close to the moment of response execution. The stop-signal paradigm has been used in a variety of research areas to examine the neural, cognitive, and developmental aspects of response inhibition in healthy as well as clinical populations (e.g., Aron & Poldrack, 2006; Badcock, Michie, Johnson, & Combrinck, 2002; Bissett & Logan, 2011; Chevalier, Chatham, & Munakata, 2014; Fillmore, Rush, & Hays, 2002; Forstmann et al., 2012; Hughes, Fulham, Johnston, & Michie, 2012; Matzke, Hughes, Badcock, Michie, & Heathcote, 2017; Schachar & Logan, 1990; Verbruggen, Stevens, & Chambers, 2014; Williams, Ponesse, Schachar, Logan, & Tannock, 1999).

Performance in the stop-signal paradigm has been conceptualized as a race between two competing processes: a go process that is triggered by the choice stimulus and a stop process that is triggered by the stop signal. If the go process wins, the response is executed; if the stop process wins, the response is inhibited (Logan, 1981; Logan & Cowan, 1984).

The stop-signal paradigm owes its popularity to the underlying race model that enables estimation of the covert latency of the stop process, known as stop-signal RT (SSRT). SSRTs can be estimated using traditional non-parametric methods or the recently developed Bayesian parametric "BEESTS" approach (for an overview, see Matzke, Verbruggen, & Logan, in press). The non-parametric approach provides researchers with a summary measure of the latency of stopping, such as mean SSRT (Logan, 1994). As is well-known in the response time literature, summary measures can mask important features of the data (e.g., Heathcote, Popiel, & Mewhort, 1991). The BEESTS approach therefore enables researchers to estimate the entire distribution of SSRTs using the assumption that go RTs and SSRTs follow an ex-Gaussian distribution (Matzke, Dolan, Logan, Brown, & Wagenmakers, 2013; Matzke, Love, et al., 2013). Despite the fact that the finishing times of the stop process cannot be directly observed, parameter recovery studies indicate that both approaches produce accurate SSRT estimates if their assumptions are met and a sufficient number of stop-signal trials are available (e.g., Band, van der Molen, & Logan, 2003; Matzke, Dolan, et al., 2013).

Using a choice go task in the stop-signal paradigm, rather than simple detection of the onset of a stimulus, has the advantage that it minimizes anticipatory responses, because accurate choices cannot be made without processing the stimulus to some degree. However, it also means that the stop-signal task mismatches the standard race model, which assumes only a single go process or "runner". A single go runner corresponds to detection, whereas to properly represent choice, a model must postulate a runner for each potential response. Despite the rich history of cognitive models to simultaneously account for RTs and choice accuracy using evidence accumulation processes (e.g., Brown & Heathcote, 2008; Ratcliff, 1978; Ratcliff, Smith, Brown, & McKoon, 2016; Ratcliff & Smith, 2004), neither the non-parametric nor the BEESTS framework addresses the choice component of the go task, and hence cannot explicitly account for "go errors". Non-parametric methods collapse correct and error RTs in a single distribution (Verbruggen, Logan, & Stevens, 2008), whereas BEESTS typically discards go errors as contaminants, and relies only on correct RTs for estimating stopping latencies (e.g., Matzke, Love, et al., 2013).

The practice of treating go errors as contaminants is not a problem if the choice is easy enough so that errors are rarely made, and if the distribution of RTs for each choice is identical. In practice, the distribution of errors is rarely checked, and even for easy choices, errors always occur for at least some participants, if only at a relatively low level (e.g., Bissett & Logan, 2011; Logan, Van Zandt, Verbruggen, & Wagenmakers, 2014; White et al., 2014). Moreover, error rates may differ among experimental manipulations, and certain clinical conditions, such as schizophrenia, may also foster error-prone performance (e.g., Hughes et al., 2012). Surprisingly, we are unaware of any previous study that investigated what error rate, or what difference in error rate, it is safe to ignore.

More broadly, the restriction to easy choices means that the standard race model may not be used to investigate response inhibition in the full range of choice tasks used in experimental psychology, which can involve a level of difficulty that results in non-negligible levels of errors, or which can rely on manipulations that affect error rates. Following the cognitive-modeling tradition, Logan et al. (2014) developed a general race model with one evidence accumulation process (runner) per choice, and applied it to data with low error rates and thousands of stop-signal trials per participant. Although this fully cognitive-process-model approach is theoretically attractive, associated estimation problems make it difficult to apply it in practice, especially with the number of stop-signal trials—rarely more than 200— typically collected in most experimental investigations.

To address this limitation, in this paper, we blend measurement and cognitive-process approaches in order to extend the standard race model to multiple response alternatives, and hence equip the model to account for errors on the go task. Our developments expand the scope of the stop-signal paradigm to the study of response inhibition in the context of difficult as well as easy choices. We show that our model has good measurement properties, and so can be practically applied in the broad range of tasks and populations studied in experimental psychology where the number of stop-signal trials that can be obtained from each participant may be limited.

We combine our treatment of go errors with two other extensions that better enable researchers to deal with stop-signal data collected in the real world. Real stop-signal data are often contaminated by the effects of processes other than inhibition, and so rarely—if ever– adhere to the idealized circumstances assumed by the standard race model. In particular, we build on the mixture-likelihood extension of BEESTS developed by Matzke, Love, and Heathcote (2017) to account for failures to launch the stop process, which we refer to as "trigger failures" (Logan, 1994). Trigger failures can occur at an elevated rate in clinical populations, and are also present more generally, albeit at a lower rate (Matzke, Hughes, et al., 2017). Here we use the same method to account for failures to launch the go process and hence errors of omission in the go task, which we refer to as "go failures". It is important to account for trigger failures because they can spuriously inflate SSRT estimates and cause deficits of attention to be mistaken for deficits of inhibition (see Matzke, Hughes, et al., 2017). Go failures, which are not uncommon in children and clinical populations (Tannock, Schachar, Carr, Chajczyk, & Logan, 1989), have the opposite effect; they masquerade as increased inhibitory ability and reduce SSRT estimates. For instance, researchers may erroneously conclude that two groups differ in SSRT because of differences in go failures that spuriously inflate the apparent inhibitory ability of one group of participants. Alternatively, go failures and trigger failures can also mask differences in SSRT estimates when biases from the different sources trade off with each other or with differences in the latency of stopping.

Our goal is to develop a flexible and unified modeling framework for the stop-signal paradigm that, for the first time, takes choice errors as well go-and trigger failures into account. We believe that our approach is an important advance in response inhibition research as it provides the first complete characterization of performance in the stop-signal paradigm. Addressing go errors and go-and trigger failures in a unified framework is essential because their combined effects on SSRT estimates are difficult to anticipate. In fact, as we demonstrate shortly, even a single contaminant in isolation can produce surprisingly strong distortions. For instance, we show that go errors—even when infrequent ($\sim 2.5\%$)—can bias parametric SSRT estimates, especially when, as is commonly the case, error responses are slower than correct responses. We also demonstrate that our unified approach ameliorates these strong distortions and interactions, regardless whether contamination occurs at a low or high level. Given that it seems unlikely—and is certainly unproven—that inhibition is a unitary construct, whereby measurements made with one easy choice task automatically generalize to other perhaps harder choice tasks, our extension of the race model to difficult choices is of general benefit to experimental psychologists. To facilitate the adoption of the developments, the software implementation of our modeling framework is available on the Open Science Framework at `https://osf.io/pbwx8`.

In what follows, we first review the standard BEESTS model (Matzke, Love, et al., 2013), which forms the basis of our modeling framework. We then develop the unified model of go failures and trigger failures, and test its estimation properties in a parameter recovery study. Next, we develop and test the full model that also incorporates choice errors. Finally, we apply the resulting model to novel stop-signal data that features manipulation of task difficulty and show that it provides relatively precise parameter estimates with only 168 stop-signal trials per participant. Importantly, we compare the performance of our framework to the standard BEESTS model with trigger failures (i.e., the model already investigated by Matzke, Love, & Heathcote, 2017) that does not account for go errors. We show that both models appear to accurately describe performance, so goodness-of-fit alone is insufficient to alert researchers to mis-specification with respect to go errors. We find non-negligible differences between the parameter estimates produced by the two models, demonstrating that ignoring go errors can cause fictitious inhibitory differences and may mislead researchers.

BEESTS: Bayesian Estimation of SSRT Distributions

The present unified framework is based on BEESTS (Matzke, Dolan, et al., 2013; Matzke, Love, et al., 2013), a Bayesian parametric race model that enables the estimation of the entire distribution of unobservable SSRTs. Following the standard race model, BEESTS assumes that response inhibition is determined by the relative finishing times of two independent processes: a stop process and a single go process. On any given trial, if go RT is slower than $SSRT + SSD$, the go RT is inhibited; if go RT is faster than $SSRT + SSD$, the go RT cannot be inhibited and results in a signal-respond RT.

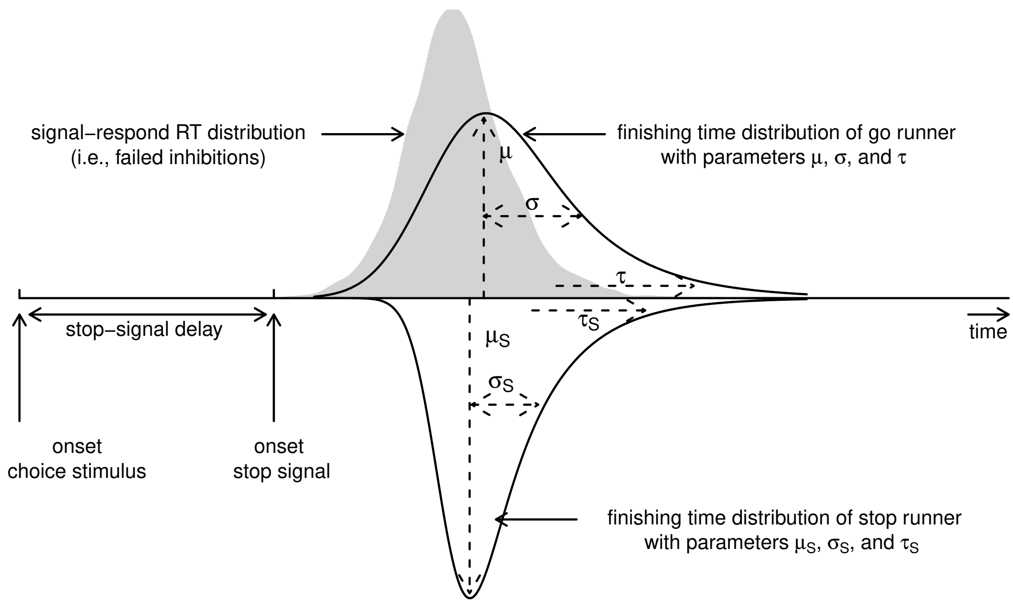

Figure 1. *Ex-Gaussian stop-signal race model with a stop runner and a single go runner.* Available at https://tinyurl.com/y86ggrj7 under CC-BY license https://creativecommons.org/licenses/by/2.0/

.

As shown in Figure 1, BEESTS assumes that go RTs and SSRTs follow an ex-Gaussian distribution. The ex-Gaussian is a frequently used descriptive RT distribution obtained by the convolution of a Gaussian and an exponential random variable (Heathcote et al., 1991; Hohle, 1965; Matzke & Wagenmakers, 2009; Ratcliff, 1978). The μ and σ parameters quantify the mean and standard deviation of the Gaussian component, and τ reflects the slow tail of the distribution. The probability density function of the ex-Gaussian distribution is:

$$f(t; \mu, \sigma, \tau) = \frac{1}{\tau} \exp\left(\frac{\mu - t}{\tau} + \frac{\sigma^2}{2\tau^2}\right) \Phi\left(\frac{t - \mu}{\sigma} - \frac{\sigma}{\tau}\right), \text{for } \sigma > 0,\, \tau > 0, \tag{1}$$

where Φ is the standard normal distribution function, defined as

$$\Phi\left(\frac{t-\mu}{\sigma}-\frac{\sigma}{\tau}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{t-\mu}{\sigma}-\frac{\sigma}{\tau}}\exp\left(\frac{-y^2}{2}\right)dy. \tag{2}$$

The distribution function of the ex-Gaussian distribution is:

$$F(t;\mu,\sigma,\tau)=\Phi\left(\frac{t-\mu}{\sigma}\right)-\exp\left(\frac{\sigma^2}{2\tau^2}-\frac{t-\mu}{\tau}\right)\Phi\left(\frac{t-\mu}{\sigma}-\frac{\sigma}{\tau}\right), \tag{3}$$

and its mean and variance equal

$$\mathrm{E}=\mu+\tau \tag{4}$$

and

$$\mathrm{Var}=\sigma^2+\tau^2, \tag{5}$$

respectively.

BEESTS relies on a separate set of ex-Gaussian parameters to describe the go RT and SSRT distributions: μ, σ, and τ for go RTs and μ_S, σ_S, and τ_S for SSRTs. Following traditional non-parametric methods, BEESTS assumes both context and stochastic independence and hence treats the go RT distribution on go trials as the underlying distribution of go RTs on stop-signal trials (Logan & Cowan, 1984).

BEESTS was developed within the Bayesian framework, partly because maximum-likelihood estimation (Myung, 2003) is computationally infeasible for the hierarchical extension of the model. BEESTS enables researchers to infer the posterior distribution of the model parameters by updating the prior distributions with incoming data. The prior distribution reflects existing knowledge about the parameter. The posterior distribution reflects knowledge about the parameter after the data have been observed. The central tendency of the posterior, such as the mean and median, may be used as a point estimate for the parameter. The 95% credible interval of the posterior (i.e., area between 2.5^{th} and 97.5^{th} percentile) encompasses the range of values that contains the true value of the parameter with 95% probability; the wider the 95% credible interval, the greater the uncertainty of the estimate. Bayesian inference is particularly suited for cognitive modeling because it offers a coherent inferential framework, which allows researchers to respect the complexity of the data-generating process and incorporate prior information (see also Lee, 2011). We provide a more elaborate explanation of the basic concepts of Bayesian inference in the Supplemental Materials (`https://osf.io/me26u/`). For comprehensive introductions to Bayesian methods in general and Bayesian cognitive modeling in particular, the reader is referred to Edwards, Lindman, and Savage (1963), Farrell and Lewandowsky (2018), Gelman and Hill (2007), Kruschke (2010), Lee and Wagenmakers (2013), and Wagenmakers et al. (2018).

A Unified Framework for Modeling Stop-Signal Data

In this section, we first extend BEESTS to simultaneously account for go failures and trigger failures. We then show how this model can be augmented to accommodate go errors. The first extension relies on a mixture-likelihood approach (e.g., Ratcliff & Tuerlinckx, 2002) to model go failures and trigger failures. The second approach adds an additional runner to the standard race model to accommodate go errors and extend the model to difficult choice tasks. Figure 2 presents an overview of the various models and show how the three-runner model with go-and trigger failures can be reduced to the standard two-runner model by dropping the additional go runner (i.e., shaded plates and gray arrows) and the go failure (P_{GF}) and trigger failure P_{TF} parameters (i.e., unshaded plates and black arrows). Models explored in the present article are described in plates with solid edges; other possible models not explored here are shown in plates with dashed edges.

We present a series of large-sample (i.e., asymptotic) parameter-recovery studies in order to verify the identifiability of the extensions and establish that the proposed model can be considered a "measurement model" in which the data-generating parameters provide the best fit to the data asymptotically (e.g., Heathcote, Brown, & Wagenmakers, 2015; Miletic, Turner, Forstmann, & van Maanen, 2017). The small-sample performance of the ex-Gaussian distribution—also in the context of the stop-signal paradigm—has been explored elsewhere (e.g., Cousineau, Brown, & Heathcote, 2004; Farrell & Ludwig, 2008; Heathcote, Brown, & Mewhort, 2002; Matzke, Love, & Heathcote, 2017; Matzke, Dolan, et al., 2013). Although these results are expected to generalize to the present approach, we urge readers to perform parameter-recovery simulations using the tutorial provided with the software, especially in the context of non-standard applications, such as stop-signal tasks embedded in recognition memory or lexical decision paradigms. Our software implementation offers a large degree of flexibility in implementing and testing paradigm-specific stop-signal models (Heathcote et al., 2018).

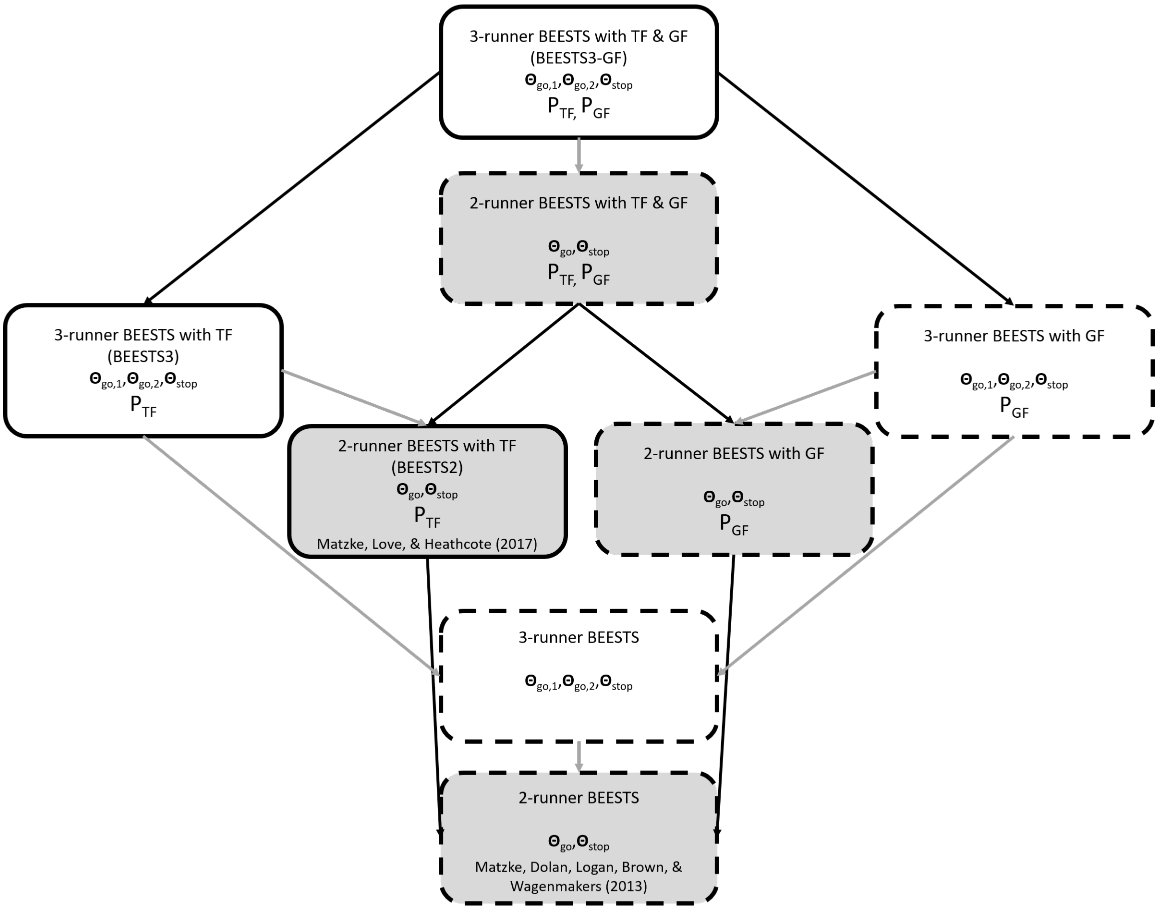

Figure 2. *Overview of the various ex-Gaussian race models.* Available at `https://tinyurl.com/y7r4m5q8` under CC-BY license `https://creativecommons.org/licenses/by/2.0/`

Modeling Go Failures and Trigger Failures

Figure 3 shows the effects of go failures and trigger failures on the inhibition function. The inhibition function, which plays a crucial role in SSRT estimation, describes the relationship between signal-respond rate and SSD. The black dots outline an inhibition function for a situation where the go and stop processes are triggered reliability on every stop-signal trial. The inhibition function increases steeply with increasing SSD and asymptotes at 0 for short and at 1 for long SSDs. The gray dots outline an inhibition function with 15% go failures; go failures decrease the steepness and the upper asymptote of the inhibition function, resulting in underestimation of stopping latencies. The gray crosses outline an inhibition function with 15% trigger failures; trigger failures decrease the steepness and increase the lower asymptote of the inhibition function, resulting in overestimation of stopping latencies. The gray triangles outline an inhibition function with 15% go and 15% trigger failures; the simultaneous presence of go and trigger failures decreases the upper and increases the lower asymptote, and further decreases the steepness of the inhibition function. As a result, the inhibition function with both types of triggering deficiencies crosses the inhibition function without go-and trigger failures (i.e., black dots).

Go failures can be accounted for in the non-parametric framework by correcting the inhibition function using the observed number of omissions on go trials without a stop signal (Tannock et al., 1989). Trigger failures cannot be accounted for by non-parametric methods, but can be straightforwardly modeled in the BEESTS framework by augmenting the standard BEESTS model with an additional parameter, P_{TF}, that quantifies the probability of trigger failures (Matzke, Love, & Heathcote, 2017). The resulting mixture model not only corrects the inhibition function but also estimates the unobservable probability that participants fail to trigger the stop process. As shown in Figure 2, we denote the trigger-failure BEESTS model as BEESTS2, where “2” stands for a race between two processes: a stop process and a go process. Despite the well-known methodological problems associated with their presence, go failures and trigger failures have not yet been modeled simultaneously in a unified framework.

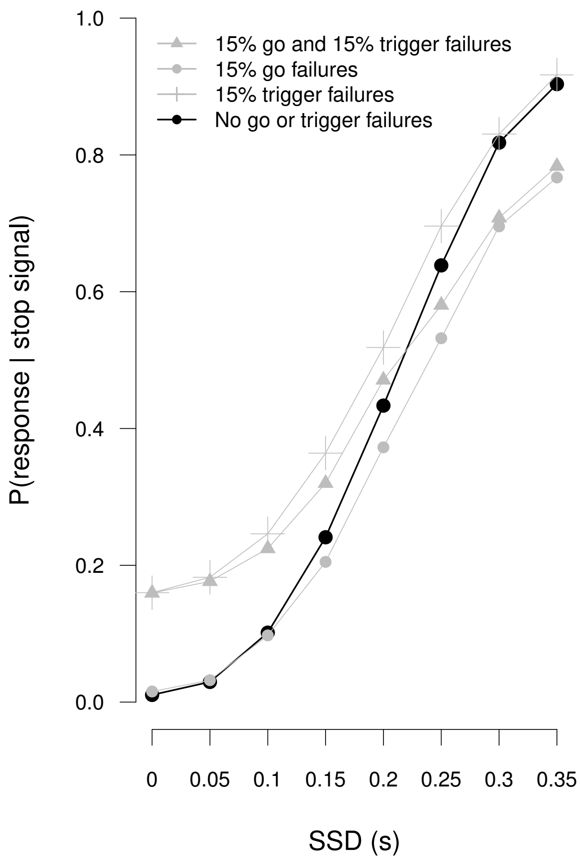

Figure 3. *The effect of go failures and trigger failures on the inhibition function.* P(response | stop signal) = signal-respond rate; SSD = stop-signal delay. The inhibition functions are based on synthetic data. Available at `https://tinyurl.com/ycosw6aj` under CC-BY license `https://creativecommons.org/licenses/by/2.0/`

To address this limitation, we augmented BEESTS2 with an additional parameter, P_{GF}, that quantifies the probability of go failures. As shown in Figure 2, we denote the resulting mixture model as BEESTS2-GF, where "GF" stands for go failures. The term *mixture* reflects the structure of the model's likelihood, which is a weighted sum of an ex-Gaussian likelihood (when there is no race because there is only one runner) and the likelihood of the minimum of two ex-Gaussians (when there are two runners).

According to BEESTS2-GF, go RTs result from go trials where the go process was successfully triggered with probability $1 - P_{GF}$. The likelihood of a response on go trial g, $g = 1, ..., G$, at time $T = t$ is then:

$$L_{GO}(\boldsymbol{\theta}_{go}, P_{GF}; t_g) = P_{GF} + (1 - P_{GF}) \times f(t_g; \boldsymbol{\theta}_{go}), \tag{6}$$

where $f(t; \boldsymbol{\theta}_{go})$ is the ex-Gaussian probability density function (Equation 1) of the finishing time distribution of the go process with parameters $\boldsymbol{\theta}_{go} = (\mu, \sigma, \tau)$.

Signal-respond RTs result from stop-signal trials where the go process was successfully triggered with probability $1 - P_{GF}$. Following Matzke, Love, and Heathcote (2017), signal-respond RTs are produced with (1) probability P_{TF} if the stop process was not triggered; or (2) probability $1 - P_{TF}$ if the stop process was triggered but finished after the go process (i.e., go RT $< SSD + SSRT$). The likelihood of a response on signal-respond trial r, $r = 1, ..., R$, at time $T = t$ is then:

$$\begin{aligned} &L_{SR}(\boldsymbol{\theta}_{go}, \boldsymbol{\theta}_{stop}, P_{TF}, P_{GF}; SSD, t_r) = \\ &\quad (1 - P_{GF}) \times \\ &\qquad \Big(P_{TF} \times f(t_r; \boldsymbol{\theta}_{go}) + \\ &\qquad (1 - P_{TF}) \times f(t_r; \boldsymbol{\theta}_{go}) \times S(t_r; \boldsymbol{\theta}_{stop}, SSD)\Big), \end{aligned} \tag{7}$$

where $S(t; \boldsymbol{\theta}_{stop})$ is the ex-Gaussian survival function of the finishing time distribution of the stop process defined as $1 - F(t; \boldsymbol{\theta}_{stop})$ (Equation 3) with parameters $\boldsymbol{\theta}_{stop} = (\mu_S, \sigma_S, \tau_S)$. P_{GF} is assumed to be independent of SSD and trial type (i.e., go, signal respond, and signal inhibit).

Successful inhibitions are produced with (1) probability $P_{GF} \times P_{TF}$ if neither the go nor the stop process was triggered; or (2) probability $P_{GF} \times (1 - P_{TF})$ if only the stop process was triggered; or (3) probability $(1 - P_{GF}) \times (1 - P_{TF})$ if both the go and the stop processes were triggered and the stop process finished before the go process (i.e., go RT $> SSRT + SSD$). The likelihood of a successful inhibition on signal-inhibit trial s, $s = 1, ..., S$, is then:

$$
\begin{aligned}
&L_S(\boldsymbol{\theta}_{go}, \boldsymbol{\theta}_{stop}, P_{TF}, P_{GF}; SSD, t_s) = \\
&\quad P_{GF} \times P_{TF} + \\
&\quad P_{GF} \times (1 - P_{TF}) \overbrace{\int_{-\infty}^{\infty} f(t_s; \boldsymbol{\theta}_{stop}, SSD) dt_s}^{1} + \\
&\quad (1 - P_{GF}) \times (1 - P_{TF}) \times \int_{-\infty}^{\infty} f(t_s; \boldsymbol{\theta}_{stop}, SSD) \times S(t_s; \boldsymbol{\theta}_{go}) dt_s,
\end{aligned} \tag{8}
$$

where $f(t; \boldsymbol{\theta}_{stop})$ is the ex-Gaussian probability density function of the finishing time distribution of the stop process and $S(t; \boldsymbol{\theta}_{go})$ is the ex-Gaussian survival function of the finishing time distribution of the go process. The integrals over t in Equation 8 reflect the fact that the finishing time of the stop process cannot be observed, so the likelihood of winning at each possible time point must be integrated (summed) to obtain the probability of stopping. Note that the first integral in Equation 8 equals one (as indicated by the bracket above it) and the second integral acts as the normalizing constant for the probability density function of the go RTs in Equation 7, ensuring that the distribution of signal-respond RTs integrates to one. Simplification results in:

$$
\begin{aligned}
&L_S(\boldsymbol{\theta}_{go}, \boldsymbol{\theta}_{stop}, P_{TF}, P_{GF}; SSD, t_s) = \\
&\quad P_{GF} + (1 - P_{GF}) \times (1 - P_{TF}) \times \int_{-\infty}^{\infty} f(t_s; \boldsymbol{\theta}_{stop}, SSD) \times S(t_s; \boldsymbol{\theta}_{go}) dt_s.
\end{aligned} \tag{9}
$$

Parameter Recovery. We generated a single stop-signal data set with $75,000$ go and $25,000$ stop-signal trials from BEESTS2-GF with $P_{TF} = 0.1$ and $P_{GF} = 0.1$. We chose to include P_{TF} in the data-generating process because trigger failures have been repeatedly shown to be an integral part of stop-signal performance in healthy as well as clinical populations (e.g., Matzke, Love, & Heathcote, 2017; Matzke, Hughes, et al., 2017; Skipppen et al., submitted; Weigard, submitted). SSD was set using the staircase-tracking procedure: SSD was increased by 0.05s after successful inhibitions and it was decreased by 0.05s after failed inhibitions, resulting in a overall signal-respond rate of approximately 0.50 (e.g., Logan, 1994).[1] The black triangles in Figure 4 show the data-generating go and stop parameters; the values are representative of estimates found in earlier applications of BEESTS2. The relatively high level of go failures is not uncommon in children or clinical populations (Tannock et al., 1989).

We fit the data set with the "true" data-generating BEESTS2-GF model as well as the misspecified BEESTS2 model that does not account for go failures. We used weakly informative uniform prior distributions for the go and stop parameters. The P_{TF} and P_{GF} parameters were assigned non-informative uniform distributions that covered the entire allowable range between 0 and 1 (see Matzke, Love, & Heathcote, 2017). We used this prior set-up for all our parameter recoveries. The exact specification of the prior distributions is available in the Supplemental Materials.

We used the Differential Evolution Markov Chain Monte Carlo (DE-MCMC; ter Braak, 2006) algorithm to sample from the posterior distribution of the parameters. The Supplemental Materials provide a more detailed explanation of MCMC-based Bayesian inference. DE-MCMC is particularly suited for obtaining posterior samples from cognitive models with highly correlated parameters (Turner, Sederberg, Brown, & Steyvers, 2013). We set the number of MCMC chains to three times the number of model parameters; for BEESTS2 we ran 21 and for BEESTS2-GF we ran 24 chains with over-dispersed start values. In order to reduce auto-correlation, we thinned each MCMC chain to retain only every 20^{th} posterior sample. During the burn-in period, we set the probability of a migration step to 5%. After burn-in, we turned off migration and performed only crossover steps until the chains converged to their stationary distribution. We assessed convergence using

[1]The recovery results generalize to stop-signal data sets with fixed SSDs.

visual inspection of the chains and univariate and multivariate proportional scale reduction factors ($\hat{R} < 1.1$; Brooks & Gelman, 1998; Gelman & Rubin, 1992). After convergence, we obtained an additional $20,000$ samples per chain; inference about the parameters was based on this final set of posterior samples. Unless indicated otherwise, we used the same sampling regime for all our analyses.

The results of the recovery study are shown in Figure 4. The black posterior distributions were computed with the true BEESTS2-GF model. BEESTS-GF's recovery was excellent; the parameters were estimated precisely (i.e., narrow posteriors) and the true values were well within the 95% credible interval of the posteriors. As expected, the newly added P_{GF} parameter was estimated very accurately because it is largely determined by the observed proportion of omission errors on go trials (Equation 6). The gray posterior distributions were computed with the misspecified BEESTS2 model that does not account for go failures. The go parameters were not biased by the presence of go failures as go failures were assumed to affect a random P_{GF} proportion of trials. In contrast, mean SSRT and the stop parameters were heavily biased by the presence of go failures; relative to BEESTS2-GF, BEESTS2 underestimated mean SSRT (i.e., $\mu_S + \tau_S$; Equation 4) and τ_S by 0.019s and 0.036s, respectively, and overestimated μ_S and σ_S by 0.017s and 0.035s, respectively.[2] Unmodeled go failures also caused bimodality in the posterior distribution of μ_S. The pattern of bias in μ_s and τ_s follows from the architecture of the race model and the intrinsic parameter correlations in the ex-Gaussian distribution. As shown in Figure 3, the presence of go failures decreases the steepness and the upper asymptote of the inhibition function, which results in underestimation of mean SSRT. This results in a very strong bias in τ_s as this parameter is largely informed by the slow tail of the observed signal-respond RT distribution, which typically features only a small number of observations (for similarly strong effects on τ_s as a result of mis-specification related to trigger failures, see Matzke, Love, & Heathcote, 2017). As a result of the strong negative correlation between the ex-Gaussian

[2] Note that we cannot compute traditional non-parametric SSRT estimates because the data were generated with $P_{TF} = 0.1$ and non-parametric methods cannot be used to estimate SSRTs in the presence of trigger failures (e.g., Band et al., 2003; Logan, 1994; Matzke, Love, & Heathcote, 2017).

μ_s and τ_s parameters, μ_s, which is relatively well constrained, compensates somewhat for the underestimation of τ_s, but this is clearly insufficient. Recovery of the P_{TF} parameter was unaffected by go failures.

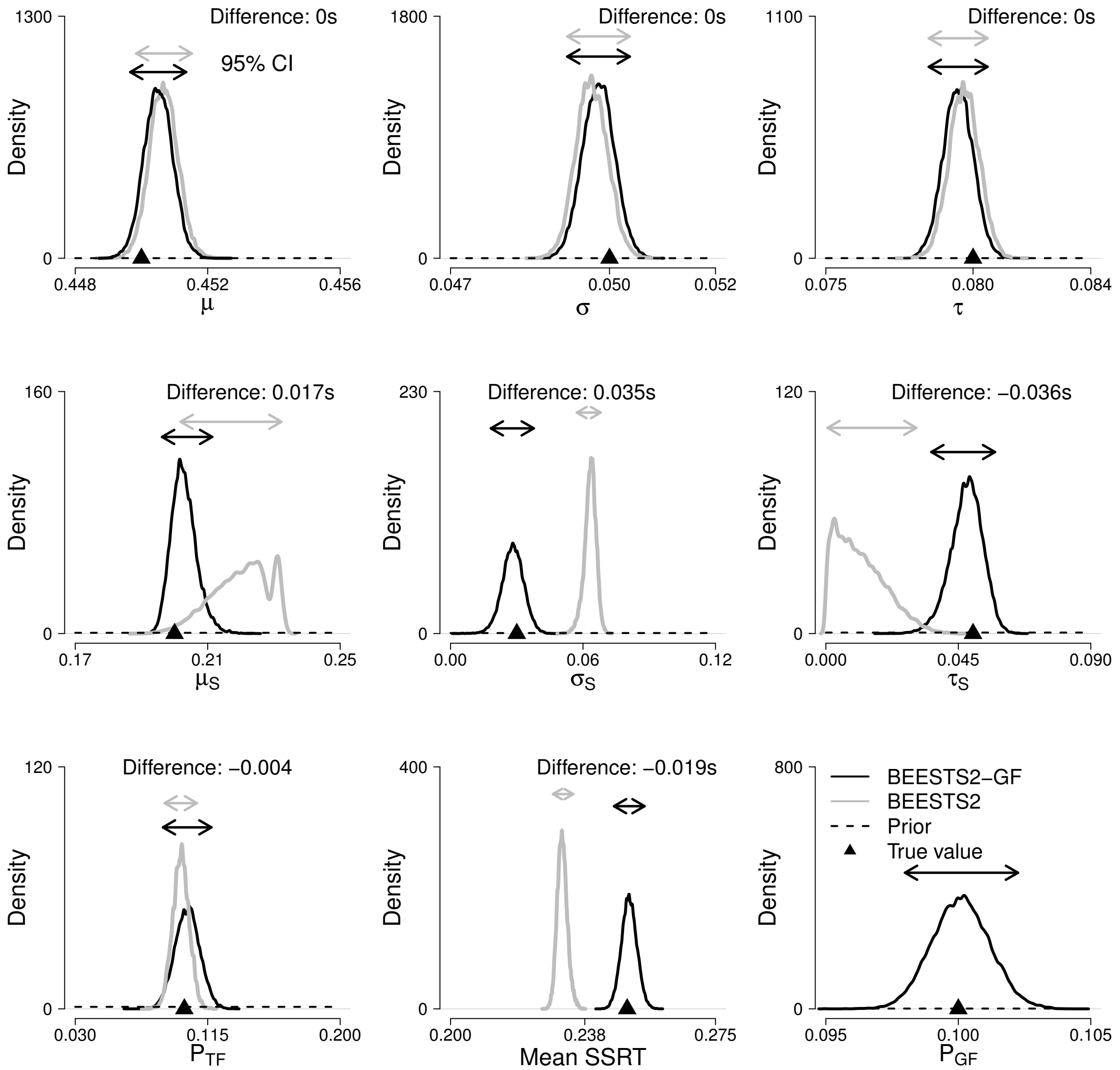

Figure 4. Bias in parameter estimates as a result of go failures. The black posterior distributions are computed with BEESTS2-GF. The gray posterior distributions are computed with the misspecified BEESTS2 that does not account for go failures. The arrows show the 95% credible interval (CI) of the posterior distributions. The dashed lines show the prior distributions. The black triangles show the true values. Mean SSRT is computed as $\mu_S + \tau_S$. Each panel shows the difference (in seconds) between the posterior mean of the BEESTS2-GF and BEESTS2 estimates.

Modeling Go Errors

In order to account for go errors and extend the model to difficult choice tasks, we augmented BEESTS2-GF with an additional go process. As shown in Figure 2, we denote this model as BEESTS3-GF, where “3” stands for a race between three independent runners, one runner that corresponds to the stop response and two runners that correspond to the two possible responses on the go task (e.g., left and right button presses). As before, we used the ex-Gaussian distribution to describe the finishing time distributions of the go and stop processes.

On a given go trial, the response and corresponding go RT is determined by the outcome of a race between the two go processes. The joint likelihood of response i on go trial g, $g = 1, ..., G$, at time $T = t$ is then:

$$L_{GO}(\boldsymbol{\theta}_{go_i}, \boldsymbol{\theta}_{go_j}, P_{GF}; t_g) = P_{GF} + (1 - P_{GF}) \times f(t_g; \boldsymbol{\theta}_{go_i}) \times S(t_g; \boldsymbol{\theta}_{go_j}), \tag{10}$$

where $f(\boldsymbol{\theta}_{go_i})$ is the Ex-Gaussian probability density function of the finishing time distribution of go process i with parameters $\boldsymbol{\theta}_{go_i} = (\mu_i, \sigma_i, \tau_i)$ and $S(\boldsymbol{\theta}_{go_j})$ is the ex-Gaussian survival function of the finishing time distribution of go process j with parameters $\boldsymbol{\theta}_{go_j} = (\mu_j, \sigma_j, \tau_j)$. The model assumes a common P_{GF} parameter for the two go processes. Note that the present approach may be extended to accommodate more than two response options on the go task (e.g., Brown & Heathcote, 2008; Heathcote & Love, 2012; Logan et al., 2014; Rouder, Province, Morey, Gomez, & Heathcote, 2015).

On a given signal-respond trial, the response and corresponding signal-respond RT is determined by the outcome of a race between the stop process and the two go processes. The joint likelihood of response i on signal-respond trial r, $r = 1, ..., R$, at time $T = t$ is then:

$$
\begin{aligned}
&L_{SR}(\boldsymbol{\theta}_{go_i}, \boldsymbol{\theta}_{go_j}, \boldsymbol{\theta}_{stop}, P_{TF}, P_{GF}; SSD, t_r) = \\
&\quad (1 - P_{GF}) \times \\
&\qquad \Big(P_{TF} \times f(t_r; \boldsymbol{\theta}_{go_i}) \times S(t_r; \boldsymbol{\theta}_{go_j}) + \\
&\qquad (1 - P_{TF}) \times f(t_r; \boldsymbol{\theta}_{go_i}) \times S(t_r; \boldsymbol{\theta}_{go_j}) \times S(t_r; \boldsymbol{\theta}_{stop}, SSD) \Big).
\end{aligned} \tag{11}
$$

Lastly, the likelihood of a successful inhibition on signal-inhibit trial s, $s = 1, ..., S$, is:

$$
\begin{aligned}
&L_S(\boldsymbol{\theta}_{go_i}, \boldsymbol{\theta}_{go_j}, \boldsymbol{\theta}_{stop}, P_{TF}, P_{GF}; SSD, t_s) = \\
&\quad P_{GF} + (1 - P_{GF})(1 - P_{TF}) \times \int_{-\infty}^{\infty} f(t_s; \boldsymbol{\theta}_{stop}, SSD) \times S(t_s; \boldsymbol{\theta}_{go_i}) \times S(t_s; \boldsymbol{\theta}_{go_j}) dt_s.
\end{aligned} \tag{12}
$$

Parameter Recovery. We assessed parameter recovery with two simulation studies. The first study focused on the three-runner BEESTS model without go failures (BEESTS3) and examined the effects of unmodeled go errors; the second study focused on BEESTS3-GF and examined the combined effects of go errors and 10% go failures. In both studies, we investigated four scenarios: low (2.5%) and high go-error (20%) rate, where errors were either on average 0.015s faster or 0.08s slower than correct responses. Fast errors typically occur when response speed is emphasized and slow errors when response accuracy is emphasized (Ratcliff & Rouder, 1998). The infrequent slow error RT condition is probably the most representative of existing stop-signal data sets.

In both studies, we generated four stop-signal data sets using staircase tracking, each with $75,000$ go and $25,000$ stop-signal trials. The black triangles in Figure 5 and Figure 6 show the data-generating go and stop parameters for BEESTS3 and BEESTS3-GF, respectively. Parameters for the go runner that matches the choice stimulus ("matching" parameters, which would be expected to have small values so that the runner finishes quickly and typically wins) are indicated by "+" subscripts, and those for the runner that mismatches the choice stimulus by "-" subscripts. We fit each data set with its respective true model as well as the misspecified BEESTS2 model that does not account for go errors and go failures. For the BEESTS2 analyses, we removed all error responses on go trials and signal-respond trials, as is typical in practice.

For brevity, we indicate the BEESTS2 parameters of the go RT distribution using the same "+" subscript as for the BEESTS3 and BEESTS3-GF matching parameters, but note that in BEESTS2 they do not correspond to the Ex-Gaussian distribution of the matching runner. Rather they largely correspond to the observed correct go RT distribution, which is a censored version of the matching distribution, with censoring corresponding to cases where the mismatching runner wins. The correspondence is not complete because these parameters also determine the censoring of the stop runner's distribution that predicts the observed signal-respond RT distribution.

Figure 5 shows the results of the first recovery study. The black horizontal lines show the 95% credible intervals of the posterior distributions computed with the true BEESTS3 model. The full posterior distributions are available in the Supplemental Materials. BEESTS3's recovery of the true values of the matching go (μ_+, σ_+, and τ_+) and stop parameters, including P_{TF}, is excellent in all four scenarios. The parameters were estimated precisely and the true values were well within the 95% credible intervals. The precision of the mismatching go estimates (μ_-, σ_-, and τ_-) was influenced by error rate and the relative speed of error and correct RTs. For slow errors, the mismatching go parameters were estimated precisely, even when error rates were low. For fast errors, the mismatching go parameters, especially τ_-, were estimated precisely only in the high-error scenario.

The gray horizontal lines in Figure 5 show the 95% credible intervals computed with the misspecified BEESTS2 model that does not account for go errors. Unmodeled go errors biased both go and stop parameters. The degree of bias varied with error rate and the relative speed of error and correct RTs. Moreover, the misspecified BEESTS2 analysis inflated the uncertainty of the stop estimates. When errors were fast and infrequent (2.5% Fast), the go and stop parameters were largely unaffected by errors. When errors were fast and frequent (20% Fast), relative to BEESTS3, BEESTS2 produced a slight 0.008s underestimation of τ_+. When errors were slow and infrequent (2.5% Slow), BEESTS2 underestimated τ_+ by 0.017s. Although the stop parameters were not affected strongly when considered in isolation, the slight downward bias in μ_S and τ_S resulted in a 0.011s underestimation of mean SSRT. Lastly, when errors were slow and frequent (20% Slow), both go and stop parameters were heavily biased. BEESTS2 underestimated τ_+ by 0.078s and overestimated μ_+ and σ_+ by 0.046 and 0.008s, respectively. Importantly, BEESTS2 resulted in a 0.031s overestimation of σ_S and a 0.031s underestimation of τ_S. This pattern resulted in a 0.025s underestimation of mean SSRT. The P_{TF} parameter was unaffected by errors.

Figure 6 shows the results of the second recovery study. The black horizontal lines show the 95% credible intervals computed with the true BEESTS3-GF model. As expected, BEESTS3-GF's recovery of the matching go and stop parameters, including P_{TF} and P_{GF}, was excellent in all four scenarios. As before, the precision of the mismatching go estimates was influenced by error rate and the relative speed of error and correct RTs.

The gray horizontal lines in Figure 6 show the 95% credible intervals computed with the misspecified BEESTS2 model that does not account for go errors and go failures. The simultaneous presence of go errors and go failures biased both go and stop parameters, including P_{TF}. As before, the misspecified BEESTS2 analysis inflated the uncertainty of the stop estimates.

As expected, the bias in the go parameters closely matched the results in Figure 5; the go parameters are influenced by go errors but not by go failures (see Figure 4). The stop parameters were heavily biased regardless of the frequency and latency of errors. Relative to BEESTS3-GF, BEESTS2 overestimated σ_S in all four scenarios, with frequent slow errors producing the largest, 0.055s, bias. In contrast, BEESTS2 underestimated τ_S, with the magnitude of the bias varying between 0.031 and 0.040s. The μ_S parameter was largely unaffected by slow errors, but was slightly overestimated in the presence of fast errors. The relatively small bias in μ_S was the consequence of the chosen parameter setting and the strong correlation between μ_S and τ_S; additional simulations confirmed that μ_S can also show a substantial downward bias dependent on the parameter setting. This pattern of bias resulted in strong underestimation of mean SSRT in all four scenarios, with frequent slow errors producing the largest, 0.037s, bias. Lastly, in contrast to the previous simulations, BEESTS2 underestimated P_{TF} by 3.2 and 4.8% in the presence of infrequent and frequent slow errors, respectively.

The recovery studies clearly demonstrated that applying the standard two-runner model to difficult choice tasks can severely bias conclusions about response inhibition. When considered in isolation, go failures did not influence the go parameters; they did however bias all three stop parameters. When considered in isolation, go errors biased both go and stop parameters, but the degree of bias varied with error rate and the relative speed of error and correct RTs. Importantly, even when infrequent, slow errors resulted in underestimation of mean SSRT. The combination of go errors and go failures resulted in heavily biased stop estimates, regardless of the frequency and latency of errors. Notably, we also observed a substantial underestimation of P_{TF}, a synergistic bias that was specific to the simultaneous presence of the two types of mis-specification.

In contrast, models that properly represented the data generating processes recovered the true values of the (matching) go and the stop parameters, including P_{TF} and P_{GF}, very well. Although the precision of the mismatching go estimates was influenced by the frequency an latency or error responses, explicitly modeling errors mitigated the bias that would have otherwise distorted SSRT estimates.

Fitting Real-World Stop-Signal Data

In this section, we illustrate the advantages of our unified modeling framework with novel stop-signal data that feature a manipulation of task difficulty. The data are available at https://osf.io/me26u/. The difficulty manipulation resulted in 12% go errors in the Easy and 25% go errors in the Difficult condition. Instructions and post-response feedback for the go task successfully emphasized the importance of responding on every trial in a fast but accurate manner, with errors that were on average 0.085s slower than correct responses and an overall go-omission rate of only 1%, although this did vary between $0 - 9\%$ over participants. The high error rate allowed us to compare the performance of BEESTS2 and BEESTS3 and demonstrate the deleterious effects of unmodeled go errors in real data. Second, the low go-omission rate allowed us to compare the performance of BEESTS3 and BEESTS3-GF and demonstrate that BEESTS3-GF may be safely used even if go omissions are infrequent, as is typical in well-motivated and trained undergraduate populations.

Task and Participants

The two-choice go task required participants to press either the “Z” or “/” keys on a standard US keyboard with their left or right index finger to indicate whether a random-dot kinematogram (RDK) displayed 45 degree left or right upward global motion, respectively. The RDK consisted of 40 dots moving in an invisible circular area of 50mm in diameter. The dots were refreshed at the rate of 30 frames per second. The coherence of the global motion was measured as the percentage of dots moving in a uniform direction, with higher coherence supporting easier perceptual judgments. On each trial, a blank screen preceded the stimulus for 0.25s, followed by a fixation point for 0.25s. The stimulus was presented for 3s.

In an initial session, participants practiced the go task over 9 blocks of 49 trials. The first block familiarized participants with the task, with difficulty gradually increased by decreasing coherence from 65% to 20%. The second and third blocks contained three levels of coherence: 5%, 10%, and 20%. Performance in these blocks determined task difficulty for the remainder of the procedure, with either 5% and 10% coherence stimuli allocated to the Difficult and Easy conditions, respectively, or 10% and 20%, depending on which pair produced an average accuracy closest to 75%. Between blocks, participants were encouraged to rest as required, and then to initiate the next block by pressing the space bar. Participants were instructed to perform quickly but accurately. Correct responses were followed by feedback on RTs; incorrect responses were followed by the feedback "Incorrect" or "Too Slow" if they failed to respond within 3s. First session results from a subset of participants used as controls were reported in Heathcote, Suraev, Curley, Gong, and Love (2015).

The stop-signal session took place on the following day and consisted of 13 blocks of 49 trials, with the first block and first trial of each block excluded from further analysis. The visual stop-signal (i.e., a gray square boarder around the go stimulus) was presented on 14 randomly selected trials per block (approximately 29% of trials). Participants were instructed to withhold their response to the go stimulus when the stop signal occurred. SSD was determined via two methods: fixed and staircase tracking. Fixed SSDs were set at 0.05s and were used for two randomly selected trails per block. The remaining 12 SSDs per block were determined using staircase tracking. The first staircase SSD in the experiment was set at 0.2s. For subsequent stop trials, successful inhibitions increased SSD by 0.033s, failed inhibitions decreased it by 0.033s.

Seventy six participants were recruited from three sources: an undergraduate student pool, the Hunter Medical Research Institute volunteer register, and the local community (see Heathcote, Suraev, et al., 2015 for details of recruitment, inclusion and exclusion criteria, and associated psychometric testing performed before the first session). Student participants received course credit and non-student participants received $40 to cover their expenses in attending the testing sessions. The study was approved by the Human Research Ethics Committee of the Newcastle University.

Due to a programming error, the maximum inter-stimulus interval for the first 15 participants was set at 2s. This resulted in truncation of the slow tail of the RT distributions for 4 participants, who were therefore excluded from analysis. Eleven participants were excluded because they responded with less than 60% accuracy on the go trials in the stop-signal session, and 6 were excluded who did not respond on greater than 3% of all go trials in both the go and stop sessions. Finally, two participants were excluded because they failed to stop on more than 75% of the stop-trials, leaving a final sample of 53 participants.

Bayesian Hierarchical Modeling

We used Bayesian hierarchical modeling to infer the posterior distribution of the BEESTS2, BEESTS3, and BEESTS3-GF parameters. Rather than estimating parameters for each participant separately, we explicitly modeled individual differences in parameter values with population-level distributions (e.g., Gelman & Hill, 2007; Lee, 2011; Matzke & Wagenmakers, 2009; Rouder, Lu, Speckman, Sun, & Jiang, 2005; Shiffrin, Lee, Kim, & Wagenmakers, 2008). The population-level distributions function as priors that "shrink" extreme participant-level estimates to the population mean. The degree of shrinkage is determined by the relative uncertainty of the estimates; uncertain estimates are pulled more strongly to the population mean than precise estimates. Bayesian hierarchical modeling can result in less variable, and on average, more accurate participant-level estimates than individual maximum likelihood or Bayesian estimation (Farrell & Ludwig, 2008; Rouder et al., 2005), especially in situations with moderate between-subject variability and scarce participant-level data (Gelman & Hill, 2007). The Supplemental Materials provide a more detailed explanation of Bayesian hierarchical modeling.

We assumed (truncated) normal population-level distributions for all model parameters. The population-level distributions were described by a set of population-level parameters: the population means and standard deviations, which were inferred from data using weakly informative priors. For instance, each participant's τ_S parameter was drawn from a normal population-level distribution truncated at 0 and 4s, with mean μ_{τ_S} and standard deviation σ_{τ_S}.[3] The population mean μ_{τ_S} was assigned a normal prior distribution truncated at 0 and 4s, with mean 0.1 and standard deviation 1. The population standard deviation σ_{τ_S} was assigned an exponential prior distribution with rate 1. The participant-level P_{TF} and P_{GF} parameters were first projected from the probability scale to the real line with a probit (i.e., standard normal cumulative distribution function) transformation before mod-

[3]The upper truncation is not necessary, but is numerically helpful.

eling them with normal population-level distributions (e.g., Matzke, Dolan, Batchelder, & Wagenmakers, 2015; Rouder, Lu, Morey, Sun, & Speckman, 2008). The exact specification of the population-level priors is available in the Supplemental Materials. Note that Bayesian parameter estimation is robust to changes in the prior as long as the data are sufficiently informative (Lee & Wagenmakers, 2013). As we demonstrate here, even with relatively diffuse priors, the type of data typically available in stop-signal studies is sufficiently informative to allow our model to provide well-behaved and relatively precise parameter estimates.

We estimated a separate set of go parameters for the Easy and the Difficult conditions. The P_{TF}, P_{GF}, and the stop parameters were constrained to be equal between the two conditions. For the BEESTS2 analysis, we removed all error responses on go and signal-respond trials. We set the number of MCMC chains to three times the number of model parameters per participant; for BEESTS2 we ran 30, for BEESTS3 we ran 48, and for BEESTS3-GF we ran 51 MCMC chains. To facilitate convergence, we first fit each participant's data separately. The mean and standard deviation of the posterior means from the individual fits were then used to obtain start values for the population means and standard deviations, respectively. The last samples from the joint posterior of the individual fits were used as start values for the participant-level parameters.

We thinned each MCMC chain to retain only every 5^{th} posterior sample. During the burn-in period, we set the probability of a migration step to 5% both at the participant and the population level. After burn-in, we performed only crossover steps until the chains converged to their stationary distribution. After convergence, we obtained an additional 100 samples per chain for BEESTS2, and BEESTS3 and 200 samples for BEESTS3-GF. Inference about the parameters was based on these final set of posterior samples.

Posterior Inference

The black horizontal lines in Figure 7 show the 95% credible interval of the posterior distributions of the population means computed with BEESTS2, BEESTS3, and BEESTS3-GF in the Easy condition; the gray lines show credible intervals computed in the Difficult condition. The triangles show the median of the posterior distributions. The population means for P_{TF} and P_{GF} were transformed back to the probability scale with a bivariate inverse-probit transformation. Bayesian p values (e.g., Klauer, 2010; Matzke, Boehm, & Vanderkerckhove, in press), computed as the proportion of posterior samples that are lower in the Difficult than in the Easy condition, are shown in the upper right corners; p values close to zero or one indicate that the posterior distribution in the Easy condition is reliably shifted to lower or higher values, respectively. The full posterior distributions of the population-level parameters and two sets of participant-level parameters are available in the Supplemental Materials.

Given the low go-omission rate, the BEESTS3 and BEESTS3-GF estimates were virtually identical. The matching go parameters, the stop parameters, and P_{TF} and P_{GF} were estimated precisely given the available data. As in the simulation study, the mismatching go parameters were estimated with quite some uncertainty, especially in the Easy condition with relatively few go errors. Bayesian p values suggested some evidence for a downward shift in the posterior distribution of τ_+ in the Easy condition for the BEESTS3 and BEESST3-GF analyses. There was no evidence for condition differences in the other go parameters.

The BEESTS2 analysis, which ignored go errors, resulted in estimates that mirrored the results of the parameter-recovery studies. In particular, relative to BEESTS3 and BEESTS3-GF, BEESTS2 underestimated the population mean of τ_+ in the Difficult condition, the condition with high error rate. This underestimation substantially decreased the overlap between the posteriors of the two difficulty conditions compared to the BEESTS3 and BEESTS3-GF analyses. Moreover, BEESTS2 inflated the uncertainty of the stop estimates, and resulted in overestimation of σ_S and underestimation of τ_S and mean SSRT. We also computed SSRT using the traditional non-parametric integration method (e.g., Verbruggen & Logan, 2009).[4] Given the negligible level of go omissions, we did not correct the estimates for go failures. The integration method resulted in an average SSRT estimate

[4]Note that traditional non-parametric SSRT methods have not been validated for the hybrid SSD procedure (staircase tracking combined with fixed short SSDs) used in our example application.

of 0.351s (SD=0.083). As non-parametric methods are known to overestimate SSRT in the presence of trigger failures, given P_{TF} of approximately 7.5%, this relatively high estimate is not unexpected and mirrors the results reported by Matzke, Love, and Heathcote (2017).

The black and gray horizontal lines in Figure 8 show the 95% credible interval of the posterior distributions of the population standard deviations computed with BEESTS2, BEESTS3, and BEESTS3-GF in the two difficulty conditions. The population standard deviation of P_{TF} and P_{GF} were transformed back to the probability scale with a bivariate inverse-probit transformation.

As before, the BEESTS3 and BEESTS3-GF estimates were virtually identical. The matching go parameters, the stop parameters, and P_{TF} and P_{GF} were estimated relatively precisely, whereas the mismatching go parameters were estimated with large uncertainty, especially in the Easy condition. There was no evidence for condition differences neither in the matching nor in the mismatching go parameters. Relative to BEESTS3 and BEESTS3-GF, the BEESTS2 analysis, which ignored go errors, underestimated individual differences in τ_+, and overestimated individual differences in σ_S and τ_S. Once again, ignoring go errors inflated the uncertainty of the stop estimates.

Goodness-of-Fit

We evaluated the absolute goodness-of-fit of the three models using posterior predictive simulations (Gelman, Meng, & Stern, 1996). We did so by comparing the observed data to predictions based on the joint posterior distributions. As we relied on the entire joint posterior to generate predictions, we not only accounted for sampling error, but also took into account the uncertainty of the parameter estimates.[5]

For each model, we randomly selected 100 parameter vectors from the joint posterior distribution of the participant-level model parameters. For each participant, we generated 100 stop-signal data sets using the chosen parameter vectors, the observed SSDs, and the

[5]We cannot formally compare the *relative* goodness-of-fit of the models because BEESTS2 accounts for only a subset of the available data: In contrast to BEESTS3-GF, BEESTS2 discards go errors and go omissions.

observed number of go and stop-signal trials. We performed three sets of posterior predictive simulations, each focusing on different aspects of the data. The first set focused on the go RT and signal-respond RT distributions, the second on inhibition functions, and the third on median signal-respond RTs. Results for the first set are provided in the Supplemental Materials. The results showed that all three models provided adequate fit to the data, even the strongly misspecified BEESTS2 model. It appears that BEESTS2's ex-Gaussian parameters can be adjusted to provide an accurate description of the observed correct go RTs even though they were not generated by an ex-Gaussian distribution (i.e., correct go RTs are generated from a *censored* ex-Gaussian distribution). Thus, evaluating goodness-of-fit is not sufficient to detect misspecification, even when error rates are high.

Inhibition Functions. The upper panels of Figure 9 show inhibition functions for BEESTS3-GF[6] (right panel) and BEESTS2 (left panel). The observed and predicted inhibition functions were averaged across participants.[7] Red bullets show the observed average signal-respond rate for each SSD-category. The gray violin plots show the distribution of the 100 predicted average signal-respond rates.

As predicted by the race model, observed signal-respond rate increased with increasing SSD. For both models, visual inspection indicated that the predictions adequately approximated the observed inhibition functions. To quantify goodness-of-fit, we computed posterior predictive p values for each SSD-category, estimated from the proportion of averaged posterior predictive samples that were greater than the data average. Extreme p values indicate that the model fails to account for the observed signal-respond rate. For BEESTS3-GF, the posterior predictive p values for the seven SSD-categories were 0.60, 0.29, 0.23, 0.06, 0.27, 1.00, and 0.77. With one exception, these p values were all in an acceptable range ($\sim 0.05 - 0.95$)[8], indicating that BEESTS3-GF provided a good description of the observed

[6]The results of the posterior predictive simulations for BEESTS3 were essentially identical to the BEESTS3-GF results, and are not presented.

[7]The SSD-categories were defined in terms of the percentiles of the SSD distribution for each participant, and then averaged over participants. This method produced an average inhibition function that better reflected the individual inhibition functions; pooling SSDs over participants before calculating the percentiles, resulted in much flatter average inhibition functions.

[8]Note that the strict cut-off of 0.05 does not apply to posterior predictive p values.

inhibition functions. For BEESTS2, the p values were 0.64, 0.76, 0.20, 0.07, 0.93, 1.00, and 1.00, indicating a similar pattern with a stronger tendency to over-predict signal-respond rate at long SSDs. However, as was the case for the CDFs, evaluating goodness-of-fit may not necessarily be sufficient to detect the misspecified nature of BEESTS2.

Signal-Respond RTs. The lower panels of Figure 9 show the results of the posterior predictive simulations using median signal-respond RT (SRRT). The observed and predicted SRRTs were averaged across participants.[9] Red bullets show the observed average SRRTs for each SSD-category. The gray violin plots show the distribution of the 100 predicted average SRRTs.

As predicted by the race model, observed SRRT increased with increasing SSD. For BEESTS3-GF, with the exception of the first and fifth SSD-category, the observed SRRTs were well within the range of predicted SRRTs. The posterior predictive p values for the seven SSD-categories were 0, 0.06, 0.91, 0.88, 0.04, 0.10, and 0.11. Similar misfit on short SSDs—SSDs that typically feature only a small number of signal-respond RTs—has been reported in numerous studies (e.g., Logan, 1981; Logan, Cowan, & Davis, 1984). These results indicate that BEESTS3-GF provided an adequate description of observed SRRTs on the majority of SSDs. For BEESTS2, the p values were 0, 0.01, 0.80, 0.74, 0.02, 0, and 0.08. In addition to the first and the fifth SSD-category, BEESTS2 also failed to account for observed SRRT in the second, fifth, and sixth SSD-category. Note, however, that the mis-specified BEESTS2 provided an adequate description of SRRTs on central SSDs, SSDs that typically contain the largest number of stop-signal trials and are therefore considered crucial in evaluating the descriptive accuracy of the race model (e.g., Matzke, Love, et al., 2013).

[9]SSDs were pooled over participants before calculating percentiles, so the same absolute SSD range was used to get median SSRTs for each participant, and then these SSRTs were averaged over participants. In contrast to inhibition functions, this method produced an average function that better reflected individual participants' functions.

Discussion

Descriptive statistical models prioritize good measurement properties, such as reliable parameter estimation, whereas cognitive-process models prioritize a veridical account of latent psychological mechanisms. Both emphases come with costs: the measurement approach may provide ambiguous inferences about latent processes (e.g., Matzke & Wagenmakers, 2009), whereas the cognitive-process approach can result in models with poorly identified parameters (e.g., Miletic et al., 2017; Schmittmann, Dolan, Raijmakers, & Batchelder, 2010). Contamination can challenge both approaches, by compromising the measurement model's ability to describe the data and provide stable parameter estimates, and by distorting the cognitive-process model's account of the latent psychological processes of interest.

Here we blended measurement and cognitive-process approaches in order to develop a flexible and unified modeling framework for the stop-signal paradigm that, for the first time, takes choice errors as well as go-and trigger failures into account. Our developments expand the scope of the stop-signal paradigm to the study of response inhibition in the context of difficult as well as easy choices. We showed that our model has good measurement properties and helps to ameliorate the surprisingly strong distortions caused by commonly occurring types of contamination in stop-signal data. The stop-signal paradigm is one of the most widely used procedures to measure the psychological construct of response inhibition (Logan & Cowan, 1984; Matzke, Verbruggen, & Logan, in press). To do so, it traditionally relies on a race model in which an inhibitory stop process or "runner" races with a runner representing the go process that produces a response. Estimation of the latency of the stop process is particularly challenging because when the inhibitory runner wins the race no response is made, and so its finishing time is never directly observed.

We focused on the impact of two processes that are often considered as sources of contamination in stop-signal data: erroneous go responses (i.e., go errors) and failures to respond to the go or stop stimulus (i.e., go failures and trigger failures, respectively). As demonstrated by our results, addressing go errors and go-and trigger failures in a unified framework is essential because their combined effects on SSRT estimates are difficult to anticipate. However, there is no qualitative distinction between contaminants and relevant psychological processes; whether a process is considered a contaminant can be a matter of perspective, and sophisticated contaminant models can even become part of the psychologically relevant part of the model (e.g., Lee, 2011; Vandekerckhove & Tuerlinckx, 2008). For instance, in some settings, failures to launch the go process (i.e., go failures) can reflect nuisance variables that occasionally interrupt task-relevant performance, whereas in an other settings, go failures can be of relevance to a psychological processes of interest (e.g., mind wandering; Cheyne, Solman, Carriere, & Smilek, 2009) or a clinical condition (e.g., hyperactivity; Tannock et al., 1989).

To address go failures, we built on the work of Matzke, Love, and Heathcote (2017), who used a mixture-likelihood approach to augment the already established BEESTS approach (Matzke, Dolan, et al., 2013) with the ability to account for failures to trigger the stop process (i.e., trigger failures). We used the same approach to account, for the first time, for go failures in a parametric model of the stop-signal paradigm. We showed that even moderate rates of go failures, similar to trigger failures, can markedly distort the primary estimate of inhibitory ability provided by the stop-signal paradigm, stop-signal RT (SSRT). Our results also showed that distortions resulting from go failures and trigger failures can be avoided by the proposed mixture-likelihood approach. Importantly, our framework has excellent measurement properties that allow go failures to be included in estimation even when they are rare.

Similar to go failures, go errors do not necessarily reflect task-irrelevant nuisance variables. In fact, in the context of standard choice RT tasks, choice errors are often considered as manifestation of the cognitive process of interest. In particular, evidence-accumulation models of choice processes (e.g., Brown & Heathcote, 2008; Ratcliff & McKoon, 2008) treat error rates and error RTs as integral parts of the data that enable identification of parameters corresponding to latent psychological processes. Unfortunately, these models introduce an assumption that makes estimation irregular, namely that the distribution of finishing times for each runner has a parameter-dependent lower bound. This irregularity potentially compromises any estimation method based on likelihoods (Cheng & Amin, 1983), and although often not a problem in standard choice RT tasks, the situation is more challenging with the partially observed data available in the stop-signal paradigm. For example, Logan et al. (2014) required each participant to perform thousands of trials to fit a cognitive-process model of the stop-signal paradigm in which each runner was modeled by an evidence-accumulation process.

To address go errors, we took a similar approach to Logan et al. (2014) in terms of cognitive architecture, with one runner for each possible choice response in the go task and one runner for the stop process, but we assumed that the finishing time for each runner is described by an ex-Gaussian distribution. The ex-Gaussian distribution is not realistic in a process sense, because, unlike the time to accumulate evidence, it is not bounded below by an unknown constant greater than zero that accounts for the time required to encode evidence from the stimulus. Instead, the ex-Gaussian distribution is a purely statistical model that aims to *describe* (as opposed to *explain*) the effects of experimental manipulations on the shape of RT distributions (Heathcote et al., 1991; Matzke & Wagenmakers, 2009). We showed that our approach, which extends the BEESTS model, has excellent measurement properties and could be successfully applied to an experiment where the number of trials performed by each participant was representative of many past applications of the stop-signal paradigm. Although the parameters of the ex-Gaussian distribution cannot give direct insights into psychological processes, they can be used to test hypotheses about cognitive architecture as long as predictions are formulated in terms of the statistical components—the parameters—of the distribution (e.g., Andrews & Heathcote, 2001; Matzke, Hughes, et al., 2017).

Note that the go process may be also modeled assuming a competition rather than a race between the response alternatives using, for instance, the diffusion decision model (DDM; Ratcliff, 1978; Ratcliff & McKoon, 2008). The DDM-approach would assume a competition between the two go responses, and the stop process would race against the go response that wins the competition. Although such hybrid models are certainly possible (e.g., White et al., 2014), we believe that the present race architecture is more general: race models can account for multiple choice alternatives in the go task, whereas the standard DDM can only deal with two choices.

Perhaps our most surprising result, and certainly the most important for applications of the standard two-runner BEESTS model, is that even low levels of go errors can compromise estimation of the distribution of the inhibitory runner, and, in particular, can cause underestimation of SSRTs. The stop-signal paradigm typically relies on an easy choice task, so error rates are low, but at least when error RTs are slower than correct RTs, such low error rates do not necessarily provide protection against the distortion of parametric SSRT estimates. Unfortunately, slow errors tend to occur when participants are trying to avoid errors (Ratcliff & Rouder, 1998), so instructions that encourage accurate responding foster exactly the conditions where the remaining errors are most problematic. Errors can be avoided altogether if the go task does not involve a choice, but instead only requires a response to the appearance of the go stimulus (i.e., a response that is not contingent on the identity of the go stimulus). However, this can lead to anticipatory responses that confound estimation of SSRT. This is because anticipatory responses effectively give the go runner a head start, and so increase the stop-signal delay by an unknown amount. Accurate knowledge of the stop-signal delay is critical to all methods of estimating SSRT.

Our results demonstrated that differences in go-error rates can confound attempts to measure inhibitory differences. For instance, spurious group differences in SSRT could arise due to differences in error rates, or, more insidiously, due to differences in the speed of error responses when error rates are equal. Further, as unmodeled errors inflate the uncertainty of the stop estimates, it will be more difficult to detect real inhibitory differences when errors are ignored. Whether these biases will lead to erroneous conclusions in a particular application depends on the (1) effect size (i.e., difference between conditions or populations); (2) differences in error rate and the latency of error RTs; (3) and the posterior uncertainty of the estimates, which is related to the number of trials per participant available for parameter estimation. Nevertheless, we showed that by explicitly modeling go errors we could remove their influence on estimates of the psychologically relevant processes, and that we could do so even when error rates were low and parameters related to the error-producing process were poorly estimated. Importantly, the latter finding indicates that the uncertainty of the error-related parameters did not propagate to the estimates of main interest, such as SSRT and trigger-failure rates. Although non-parametric SSRT estimation based on the standard two-runner model is relatively robust to go errors, our parametric modeling framework provides a richer characterization of stop-signal performance in terms of the probability of each response and the associated RT distributions. This in turn allows researchers to derive more specific predictions and design more stringent empirical evaluations of theories of response inhibition.

Most importantly, our empirical example demonstrated that our approach can be successfully applied to stop-signal tasks with high error rates. The standard race model (Logan & Cowan, 1984) assumes a race between a stop process and a single go process. However, to properly represent the choice embedded in the go task, a model must postulate a runner for each response option. Our empirical results clearly showed that applying the parametric two-runner model to tasks involving difficult choices is not only theoretically, but also practically, unjustified.

Our generalization of the standard race model to multiple response alternatives enables researchers, for the first time, to use the stop-signal paradigm to investigate in relatively small samples the ability to inhibit difficult as well as easy choices. In this way our modeling framework extends the applicability of the stop-signal procedure to research areas in experimental psychology, such as recognition memory, that often rely on difficult choice task and manipulations that affect error rates (e.g., Kim, Potter, Craigmile, Peruggia, & Van Zandt, 2017). Moreover, our approach enables researchers to investigate whether conclusions about response inhibition derived from easy choice tasks generalize to more difficult choices that pervade, and which are critical to effective functioning, in daily life.

Acknowledgements

DM is supported by a Veni grant (45115010) from the Netherlands Organization of Scientific Research (NWO). SC is supported by a PhD Scholarship from the Centre for Brain and Mental Health, University of Newcastle. AH is supported by an ARC Discovery project (DP120102907) and an ABC-VIP grant from the University of Amsterdam.

Data Availability Statement

The data from the example application are available on the Open Science Framework at https://osf.io/me26u/.

References

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 514–544.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: Role of the subthalamic nucleus. *Journal of Neuroscience*, *26*, 2424–2433.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*, 177–185.

Badcock, J. C., Michie, P., Johnson, L., & Combrinck, J. (2002). Acts of control in schizophrenia: Dissociating the components of inhibition. *Psychological Medicine*, *32*, 287–297.

Band, G. P. H., van der Molen, M. W., & Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychologica*, *112*, 105–142.

Bissett, P. G., & Logan, G. D. (2011). Balancing cognitive demands: Control adjustments in the stop-signal paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 392–404.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

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

Cheng, R. C. H., & Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B*, *45*, 394–403.

Chevalier, N., Chatham, C. H., & Munakata, Y. (2014). The practice of going helps children to stop: The importance of context monitoring in inhibitory control. *Journal of Experimental Psychology: General*, *143*, 959–965.

Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009, April). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*, 98–113.

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods*, *36*, 742–756.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*, 193–242.

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior.* Cambridge University Press.

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*, 1209–1217.

Fillmore, M. T., Rush, C. R., & Hays, L. (2002). Acute effects of oral cocaine on inhibitory control of behavior in humans. *Drug and Alcohol Dependence*, *67*, 157–167.

Forstmann, B. U., Keuken, M. C., Jahfari, S., Bazin, P.-L., Neumann, J., Schäfer, A., … Turner, R. (2012). Cortico-subthalamic white matter tract strength predicts interindividual efficacy in stopping a motor response. *Neuroimage*, *60*, 370–375.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge, UK: Cambridge University Press.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–807.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, *7*, 457–472.

Heathcote, A., Brown, S., & Mewhort, D. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*, 394–401.

Heathcote, A., Brown, S., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In B. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 25–48). New York: Springer.

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2018). Dynamic models of choice. *Behavior Research Methods.*

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Cognitive Science*, *3*, 292.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340–347.

Heathcote, A., Suraev, A., Curley, S., Gong, Q., & Love, J. (2015). Decision processes and the slowing of simple choices in schizophrenia. *Journal of Abnormal Psychology*, *124*, 961–974.

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*, 382-386.

Hughes, M. E., Fulham, W. R., Johnston, P. J., & Michie, P. T. (2012). Stop-signal response inhibition in schizophrenia: Behavioural, event-related potential and functional neuroimaging data. *Biological Psychology*, *89*, 220–231.

Kim, S., Potter, K., Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2017, March). A Bayesian race model for recognition memory. *Journal of the American Statistical Association*, *112*, 77–91.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*, 70–98.

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 658–676.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*, 1–7.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian modeling for cognitive science: A practical course.* Cambridge, UK: Cambridge University Press.

Logan, G. D. (1981). Attention, automaticity, and the ability to stop a speeded choice respons. In J. B. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 205–222). Hillsdale, NJ: Erlbaurn.

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 189–240). San Diego, CA: Academic Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.

Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: A model and a method. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 276-291.

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

Matzke, D., Boehm, U., & Vanderkerckhove, J. (in press). Bayesian inference for psychology. Part III: Parameter estimation in nonstandard models. *Psychonomic Bulletin & Review.*

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*, 205-235.

Matzke, D., Dolan, C. V., Logan, G. D., Brown, S. D., & Wagenmakers, E.-J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. *Journal of Experimental Psychology: General*, *142*, 1047-1073.

Matzke, D., Hughes, M., Badcock, J. C., Michie, P., & Heathcote, A. (2017). Failures of cognitive control or attention? The case of stop-signal deficits in schizophrenia. *Attention, Perception, & Psychophysics*, *79*, 1078-1086.

Matzke, D., Love, J., & Heathcote, A. (2017). A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. *Behavior Research Methods*, *49*, 267-281.

Matzke, D., Love, J., Wiecki, T. V., Brown, S. D., Logan, G. D., & Wagenmakers, E.-J. (2013). Release the BEESTS: Bayesian estimation of ex-Gaussian stop-signal reaction time distributions. *Frontiers in Quantitative Psychology and Measurement*, *4:918*, doi: 10.3389/fpsyg.2013.00918.

Matzke, D., Verbruggen, F., & Logan, G. (in press). The stop-signal paradigm. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience.* John Wiley & Sons, Inc.

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.

Miletic, S., Turner, B., Forstmann, B. U., & van Maanen, L. (2017). Parameter recovery for the Leaky Competitive Accumulator model. *Journal of Mathematical Psychology*, *76*, 25–50.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

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

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

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*, 260–281.

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.

Ridderinkhof, K. R., Van Den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*, 129–140.

Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General*, *137*, 370–389.

Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, *12*, 195–223.

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*, 491–513.

Schachar, R., & Logan, G. D. (1990). Impulsivity and inhibitory control in normal development and childhood psychopathology. *Developmental Psychology*, *26*, 710–720.

Schmittmann, V. D., Dolan, C. V., Raijmakers, M. E., & Batchelder, W. H. (2010). Parameter identification in multinomial processing tree models. *Behavior Research Methods*, *42*, 836–846.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Skipppen, P., Matzke, D., Heathcote, A., Fulham, W., Michie, P., & Karayanidis, F. (submitted). Improved estimation of response inhibition in the stop-signal task does not improve predictions of impulsivity.

Tannock, R., Schachar, R. J., Carr, R. P., Chajczyk, D., & Logan, G. D. (1989). Effects of methylphenidate on inhibitory control in hyperactive children. *Journal of Abnormal Child Psychology*, *17*, 473–491.

ter Braak, C. J. F. (2006). A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, *16*, 239-249.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*, 368–384.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, *40*, 61–72.

Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience & Biobehavioral Reviews*, *33*, 647–661.

Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP–IT: Windows executable software for the stop-signal paradigm. *Behavior Research Methods*, *40*, 479–483.

Verbruggen, F., Stevens, T., & Chambers, C. D. (2014). Proactive and reactive stopping when distracted: An attentional account. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1295.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... others (2018). Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*, 35–57.

Weigard, H. A. M. D. . H.-P. C., A. (submitted). A bayesian model-based analysis of stop-signal task performance reveals attentional causes of apparent inhibition deficits in adhd.

White, C. N., Congdon, E., Mumford, J. A., Karlsgodt, K. H., Sabb, F. W., Freimer, N. B., ... Poldrack, R. A. (2014). Decomposing decision components in the stop-signal task: A model-based approach to individual differences in inhibitory control. *Journal of Cognitive Neuroscience*, *26*, 1601–1614.

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology*, *35*, 205–213.

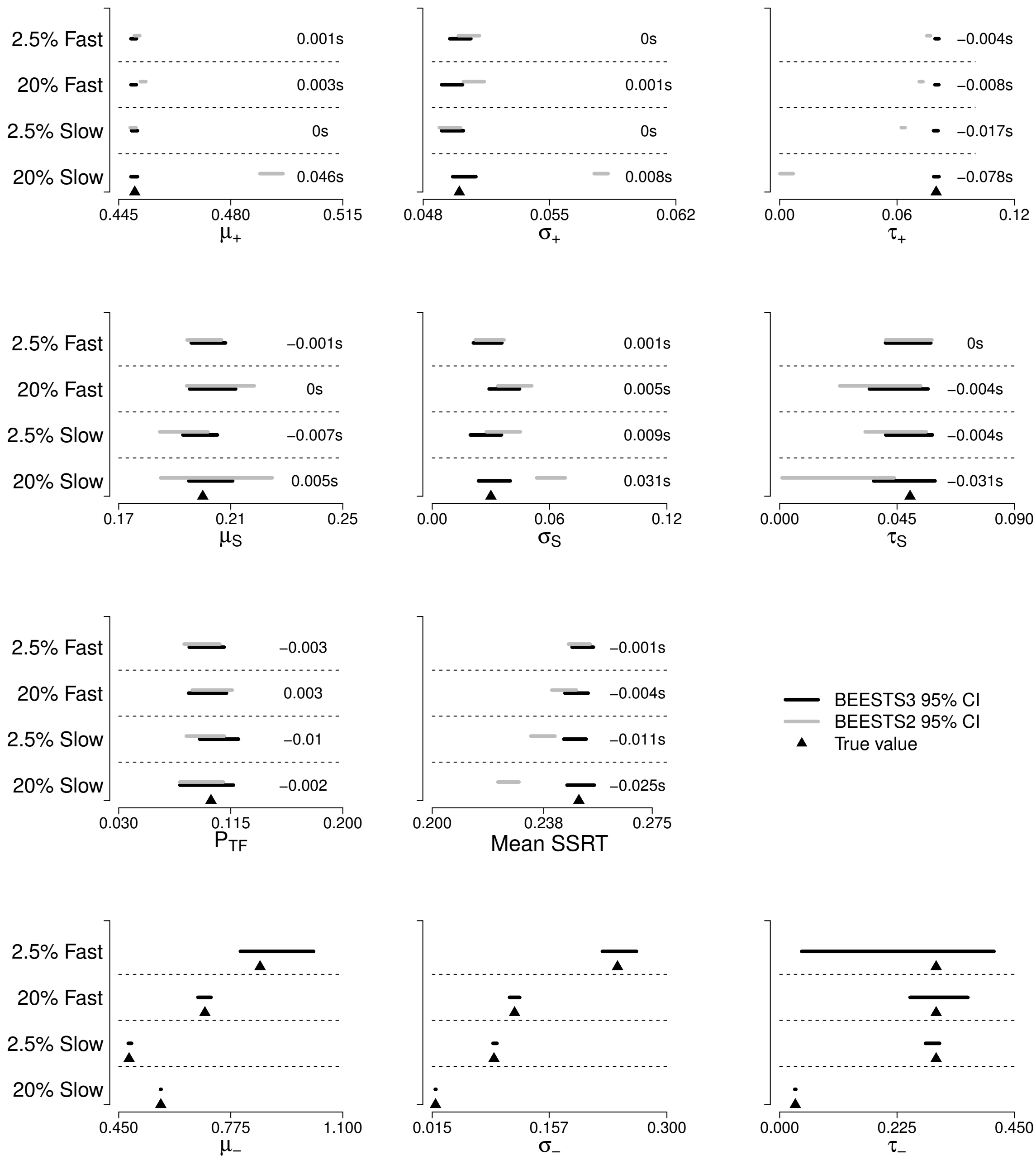

Figure 5. Bias in parameter estimates as a result of go errors. The black horizontal lines show the 95% credible intervals (CI) of the posterior distributions computed with BEESTS3. The gray horizontal lines show the 95% CIs of the posterior distributions computed with the misspecified BEESTS2 that does not account for go errors. The black triangles show the true values. The difference (in seconds) between the posterior means of the BEESTS2 and BEESTS3 estimates are shown in the upper right corners. Mean SSRT is computed as $\mu_S + \tau_S$. The subscripts + and − denote the matching and mismatching go runners, respectively.

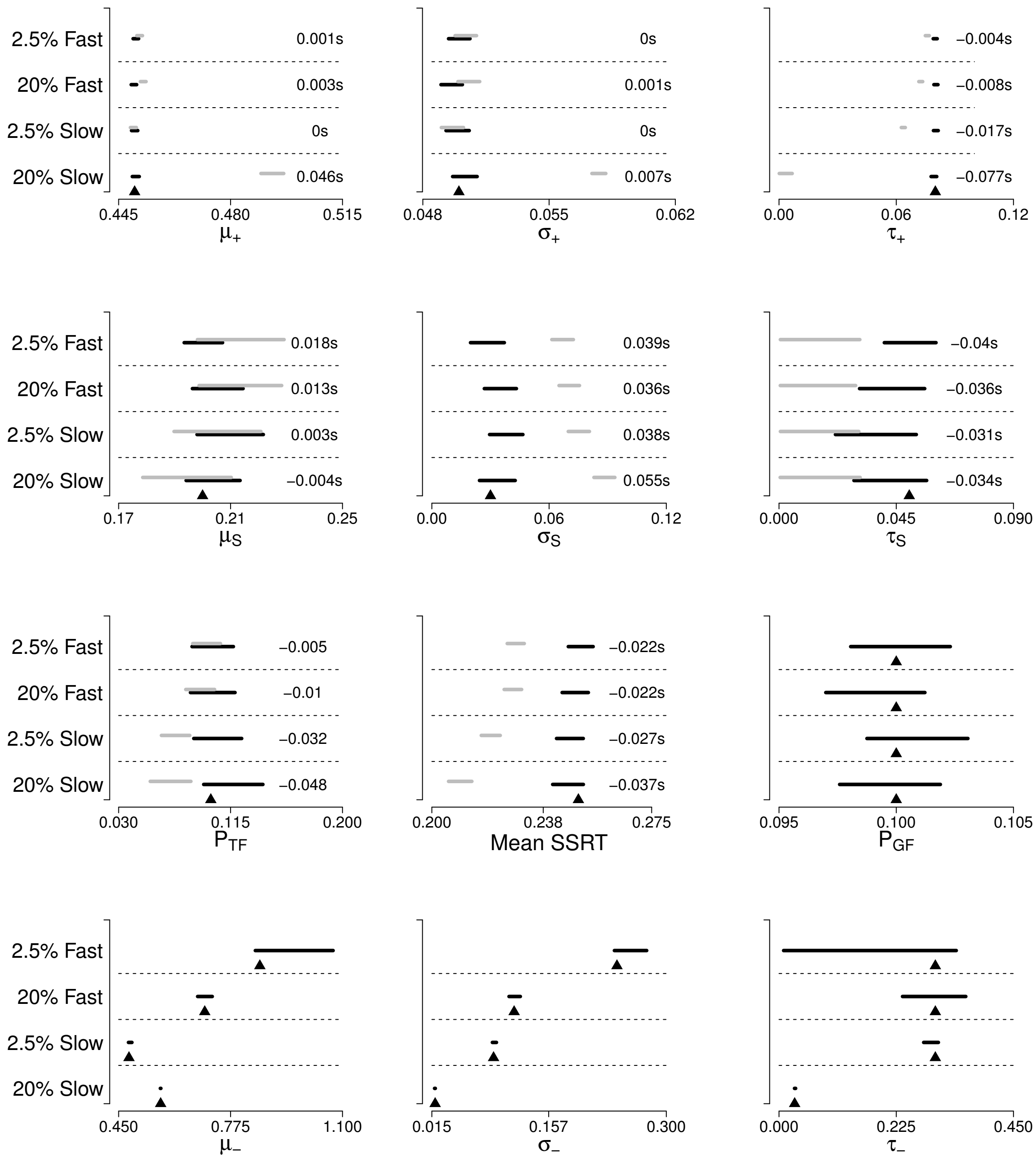

Figure 6. *Bias in parameter estimates as a result of go failures and go errors.* The black horizontal lines show the 95% credible intervals (CI) of the posterior distributions computed with BEESTS3-GF. The gray horizontal lines show the 95% CIs of the posterior distributions computed with the misspecified BEESTS2 that does not account for go errors and go failures. The black triangles show the true values. The difference (in seconds) between the posterior means of the BEESTS2 and BEESTS3-GF estimates are shown in the upper right corners. Mean SSRT is computed as $\mu_S + \tau_S$. The subscripts + and − denote the matching and mismatching runners, respectively.

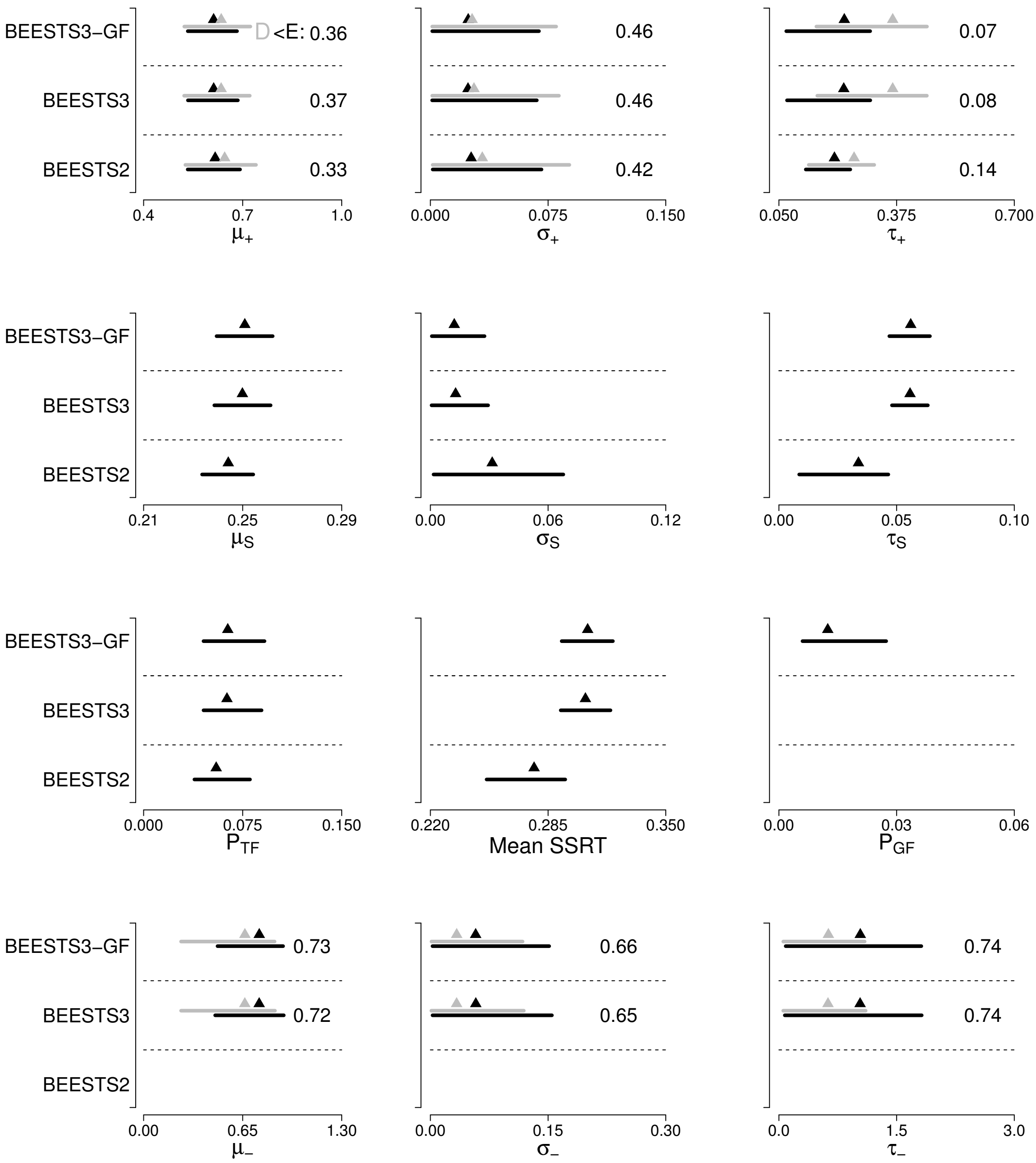

Figure 7. 95% credible intervals of the population means for the empirical data obtained with BEESTS2, BEESTS3, and BEESTS3-GF. The black horizontal lines show the 95% credible intervals (CI) of the posterior distributions in the Easy condition; the gray horizontal lines show the 95% CIs in the Difficult condition. The triangles show the posterior medians. Bayesian *p* values, computed as the proportion of posterior samples that are lower in the Difficult (D) than in the Easy (E) condition, are shown in the upper right corners. Mean SSRT is computed as $\mu_S + \tau_S$. The subscripts + and − denote the matching and mismatching runners, respectively.

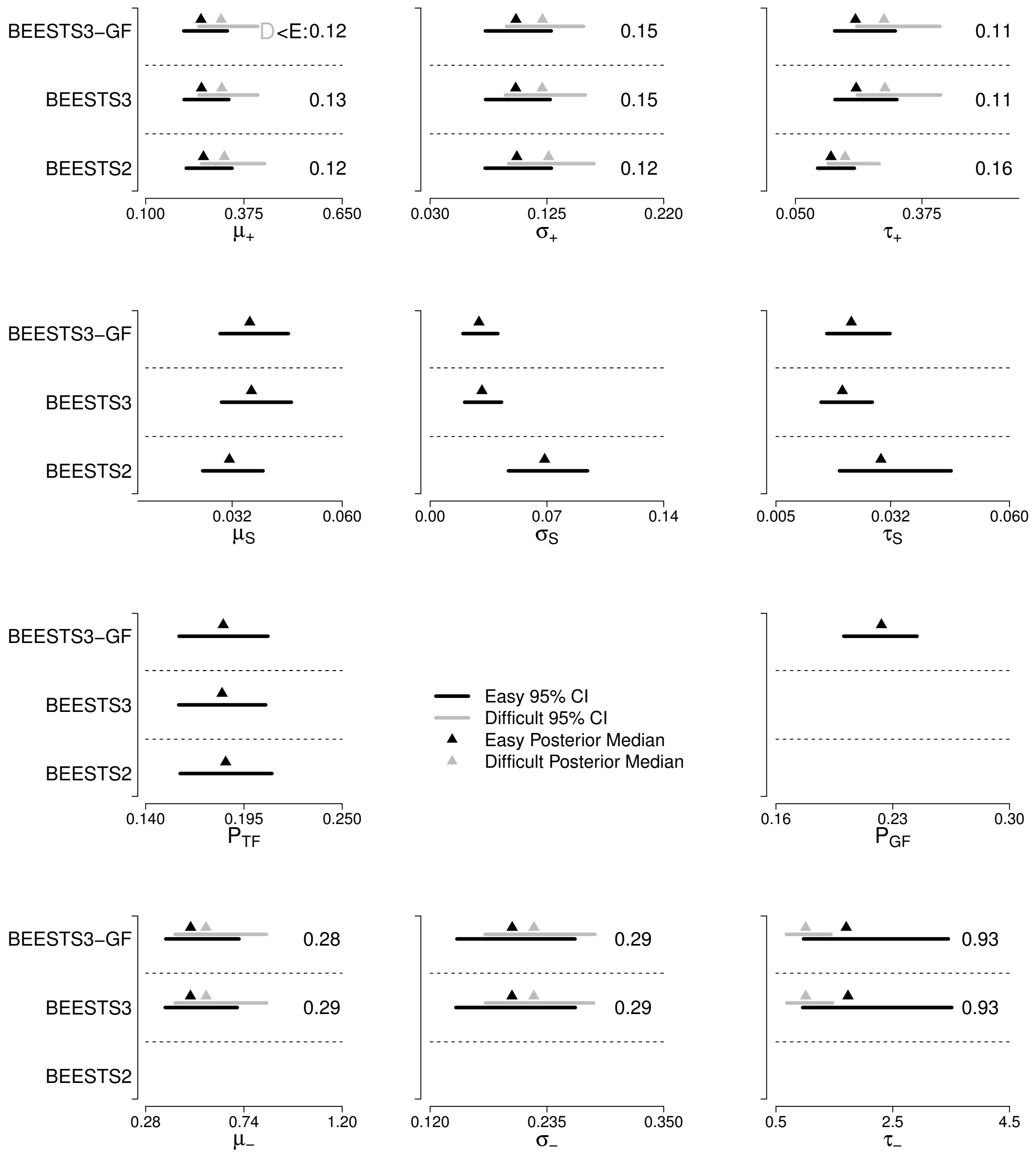

Figure 8. *95% credible intervals of the population standard deviations in the empirical data obtained with BEESTS2, BEESTS3, and BEESTS3-GF.* The black horizontal lines show the 95% credible intervals (CI) of the posterior distributions in the Easy condition; the gray horizontal lines show the 95% CIs in the Difficult condition. The triangles show the posterior medians. Bayesian *p* values, computed as the proportion of posterior samples that are lower in the Difficult (D) than in the Easy (E) condition, are shown in the upper right corners. The subscripts + and − denote the matching and mismatching runners, respectively.

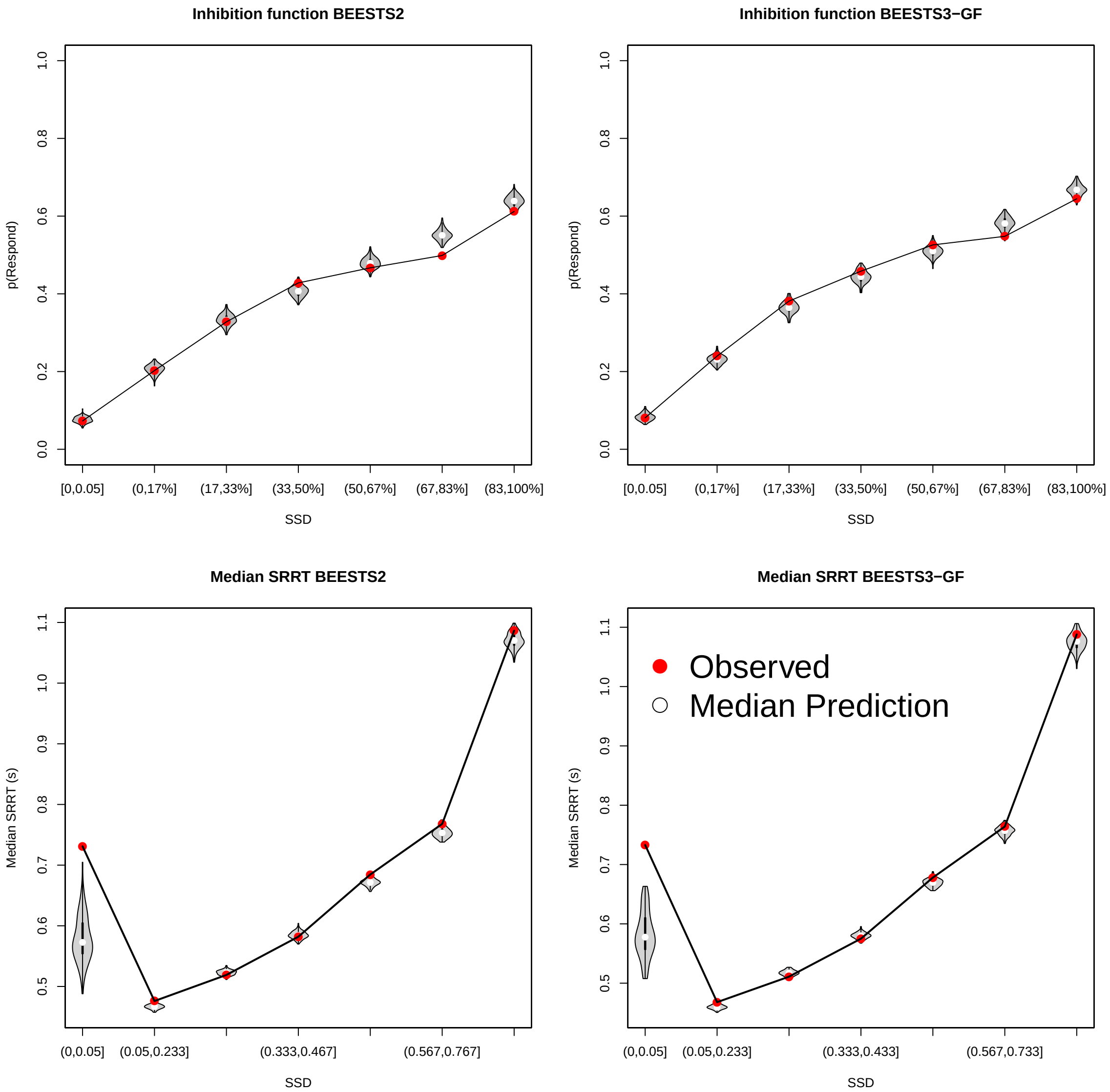

Figure 9. *Observed vs. predicted inhibition functions (top panels) and median signal-respond RTs for BEESTS2 and BEESTS3-GF.* In the top panels, red bullets show the observed average signal-respond rate ($p(Respond)$) for each SSD-category. In the bottom panels, red bullets show the observed average median signal-respond RT (SRRT) for each SSD-category. The gray violin plots show the distribution of the 100 average signal-respond rates and median SRRTs predicted by the models. The black boxplot in each violin plot ranges from the 25^{th} to the 75^{th} percentile of the predictions; the white circle represents the median of the predictions.