A cognitive process modeling framework for the ABCD study stop-signal task

Weigard, A.; Matzke, D.; Tanis, C.; Heathcote, A.

DOI
10.1016/j.dcn.2022.101191

Publication date
2023

Document Version
Final published version

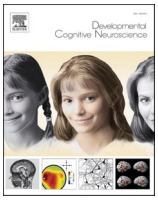
Published in
Developmental Cognitive Neuroscience

Developmental Cognitive Neuroscience 59 (2023) 101191

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

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



A cognitive process modeling framework for the ABCD study stop-signal task

Alexander Weigard [a,*], Dora Matzke [b], Charlotte Tanis [b], Andrew Heathcote [b,c]

[a] Department of Psychiatry, University of Michigan, USA
[b] Department of Psychology, University of Amsterdam, the Netherlands
[c] School of Psychology, the University of Newcastle, Australia

ARTICLE INFO

Keywords:
Inhibition
Evidence accumulation
Trigger failure
Bayesian cognitive modeling
Parameter recovery

ABSTRACT

The Adolescent Brain Cognitive Development (ABCD) Study is a longitudinal neuroimaging study of unprecedented scale that is in the process of following over 11,000 youth from middle childhood though age 20. However, a design feature of the study's stop-signal task violates "context independence", an assumption critical to current non-parametric methods for estimating stop-signal reaction time (SSRT), a key measure of inhibitory ability in the study. This has led some experts to call for the task to be changed and for previously collected data to be used with caution. We present a cognitive process modeling framework, the RDEX-ABCD model, that provides a parsimonious explanation for the impact of this design feature on "go" stimulus processing and successfully accounts for key behavioral trends in the ABCD data. Simulation studies using this model suggest that failing to account for the context independence violations in the ABCD design can lead to erroneous inferences in several realistic scenarios. However, we demonstrate that RDEX-ABCD effectively addresses these violations and can be used to accurately measure SSRT along with an array of additional mechanistic parameters of interest (e.g., attention to the stop signal, cognitive efficiency), advancing investigators' ability to draw valid and nuanced inferences from ABCD data.

Availability of data and materials: Data from the ABCD Study are available through the NIH Data Archive (NDA): nda.nih.gov/abcd. Code for all analyses featured in this study is openly available on the Open Science Framework (OSF): osf.io/2h8a7/.

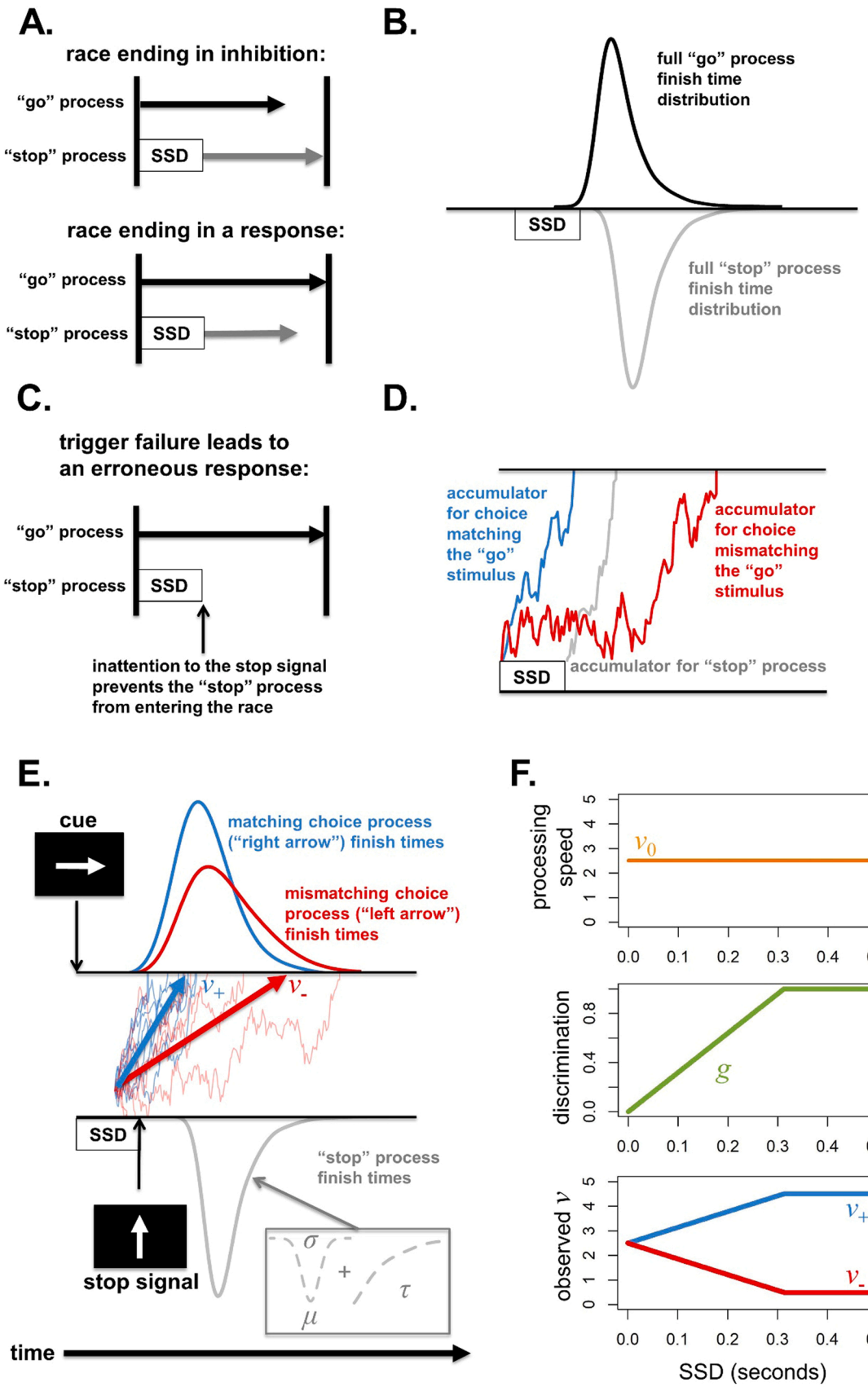
1. Introduction

Response inhibition, the ability to stop prepotent responses or actions that are no longer contextually appropriate, is a foundational component of self-control (Miyake et al., 2000; Ridderinkhof et al., 2004; Verbruggen et al., 2014) that is of key interest in research on multiple clinical conditions, including Attention-Deficit/Hyperactivity Disorder (ADHD) and problematic substance use (Boonstra et al., 2010; Gorenstein and Newman, 1980; Mahmood et al., 2013; Nigg, 2017). The stop-signal paradigm (Logan et al., 1984; Verbruggen and Logan, 2008), one of the most widely employed laboratory measures of response inhibition, probes this ability in the context of a two-choice decision task. On a minority of "stop" trials a visual or auditory "stop signal" indicating that participants must withhold their response is presented after a variable delay following the onset of the choice stimulus (the "stop-signal delay" or SSD). Depending on both the SSD and the participant's inhibitory ability, either inhibition fails, and a choice response is made (a "signal response"), or the response is successfully withheld.

An appealing feature of the stop-signal paradigm is that it was explicitly designed with a process model in mind that can be leveraged to precisely measure inhibitory ability: the "independent race model" (Logan, 1994; Logan et al., 1984; Logan and Cowan, 1984). This model posits that, on stop trials, a "go" process triggered by the choice stimulus races a "stop" process triggered by the stop signal. When the stop process wins the response is inhibited, whereas inhibition fails when the go process finishes first (Fig. 1A). By making only limited assumptions about the distributions of finishing times for these processes, including a "context independence" assumption that the go process is the same regardless of whether or not there is a stop signal, the speed of the latent stop process, or "stop-signal reaction time" (SSRT), can be easily estimated (Logan, 1994; Matzke et al., 2018). Such "non-parametric" SSRT

* Correspondence to: University of Michigan Department of Psychiatry, Rachel Upjohn Building, 4250 Plymouth Road, Ann Arbor, MI 48109, USA.
E-mail address: asweigar@med.umich.edu (A. Weigard).

https://doi.org/10.1016/j.dcn.2022.101191
Received 26 October 2022; Received in revised form 7 December 2022; Accepted 21 December 2022
Available online 22 December 2022



(caption on next page)

Fig. 1. Models of the stop signal task. (A) The original independent race model, consisting of a race between the go and stop processes while accounting for the stop signal delay (SSD). The stop process wins the first race shown, resulting in a successful inhibition, while the go process wins the second race, resulting in an incorrect "signal response". (B) The Bayesian parametric estimation of stop-signal reaction time distributions (BEESTS: Matzke et al., 2013) framework models the entire distributions of go and stop process finishing times as ex-Gaussian (i.e., right-skewed normal) distributions. (C) An example of "trigger failure" leading to a signal response by preventing the stop process from entering the race. (D) The racing-diffusion (Logan et al., 2014) model's explanation of an example stop trial; accumulators gather noisy evidence over time for the choices matching and mismatching the go stimulus as well as for the stop process. In this case, the accumulator for the matching go choice process reaches threshold before the stop process accumulator, causing a signal response to be made. (E) The "hybrid" racing-diffusion ex-Gaussian (RDEX) model framework (Tanis et al., 2022) that combines an evidence-accumulation model of the go process with an ex-Gaussian model of stop process finishing times. Go RTs result from a race between accumulators that gather noisy evidence for the choices matching and mismatching the stimulus (in this example, a right-facing arrow) at average rates of v_+ and v_-, respectively, until one accumulator crosses a response threshold. Stop process finishing times are drawn from a Gaussian distribution specified by mean (μ) and standard deviation (σ) parameters and convolved with an exponential distribution with mean τ. (F) The RDEX-ABCD model's explanation for the impact of context independence violations on stop trials. Evidence signals for the matching and mismatching accumulators on stop trials of a given SSD are the sum of a *processing speed* component that drives evidence accumulation for both choices equally and a *discrimination* component that favors the choice matching the presented stimulus. Processing speed is determined by parameter v_0 and is identical across all SSDs. The discrimination component is completely absent (equal to 0) at a 0 s SSD, as the choice stimulus is not presented, but increases linearly at the same rate g for both matching and mismatching components until they reach the level of v_+ and v_-on go trials. Therefore, the match and mismatch rates are identical on 0 s SSD trials and gradually move apart from each other at longer SSDs until they become equal to their go trial levels.

estimates are used in many neuroscientific and clinical applications to measure individuals' response inhibition (Aron and Poldrack, 2006; Lipszyc and Schachar, 2010; Nigg et al., 2006).

Given this task's popularity and the ability of SSRT to precisely index response inhibition, it is not surprising that the task has been included in one of the most ambitious research efforts of our time: the Adolescent Brain Cognitive Development (ABCD) Study (Casey et al., 2018; Garavan et al., 2018). ABCD is a multi-site collaboration in the United States that has recruited a diverse sample of over 11,000 9- and 10-year-old children and aims to follow them prospectively through at least age 20 to acquire a rich array of longitudinal data from neuroimaging, cognitive, personality, psychiatric and sociocultural domains. Although initially conceived with the goal of assessing the impact of substance use on adolescents' brains (Volkow et al., 2018), ABCD has grown into an unprecedented interdisciplinary collaboration and open data source that is beginning to drive new insights in areas as diverse as network neuroscience (Marek et al., 2019; Sripada et al., 2019), child psychopathology (Clark et al., 2021; Funkhouser et al., 2020; Mennies et al., 2020), and bilingualism (Dick et al., 2019). As several lines of work now underscore the importance of large, population-based samples for bolstering the reproducibility of behavioral and neuroscientific research (Button et al., 2013; Etz and Vandekerckhove, 2016; Falk et al., 2013; Loken and Gelman, 2017; Marek et al., 2022; Poldrack and Gorgolewski, 2014), the ABCD Study presents a critical opportunity for scientists to characterize the clinical and neural correlates of response inhibition and related neurocognitive functions.

However, concern has recently been raised that the design of the ABCD stop-signal task invalidates non-parametric SSRT estimates (Bissett et al., 2021). Specifically, the ABCD task's visual stop signal effectively masks the "go" choice stimulus, violating the independent race model's "context independence" assumption that the go choice process has identical properties across "go" and "stop" trials. As violations of this assumption can bias non-parametric estimates of SSRT, Bissett et al. (2021) have suggested that SSRT estimates from this task be interpreted with caution and have called for context independence violations in ABCD to be addressed either through novel cognitive models or changes in the task design.

In this article, we develop such a novel model and show that it both describes the impact of the masking effect and mitigates its impact on substantive inferences about response inhibition. We begin by outlining key principles of cognitive process modeling and describing previous models of the stop signal task in the context of these principles. Next, we describe the context independence violation caused by the ABCD task design and motivate our modeling solution. We then apply our solution to empirical data from the ABCD study to evaluate its ability to describe key trends, elaborate its advantages relative to alternate models, and demonstrate its utility for drawing inferences about mechanisms of neurocognitive performance in ABCD.

1.1. Key principles of cognitive process modeling

Cognitive process modeling is the practice of specifying and testing formal (i.e., mathematically and/or computationally precise) models describing psychological mechanisms that give rise to observed behavior (Heathcote et al., 2015). Beyond providing clear and testable predictions about empirical data, one of the most valuable features of cognitive process models is that they can allow the mechanistic processes they propose to be quantifiably measured. The original independent race model of the stop-signal paradigm is a simple exemplar. By positing that the outcome of response inhibition is a function of a race between the latent go and stop processes, the latency of the latter process (SSRT) provides an index of an individual's ability to inhibit responses. Widely used "non-parametric" estimation procedures enable summary measures of SSRT to be inferred from simple behavioral statistics (e.g., mean RT on "go" trials and mean SSD) (Verbruggen et al., 2019). In contrast, by employing specific parametric assumptions about the go and stop processes, the model-fitting approaches we focus on below provide a more flexible and complete account of the full distributions of behavioral data (i.e., go RT and SSRT distributions).

Although an extensive review of cognitive modeling practices is beyond the scope of this article (instead, see: Heathcote et al., 2015), several principles have critical relevance when considering models of the ABCD stop-signal task. First, for a cognitive model's estimates of latent quantities to be useful for making inferences, the model itself must provide an adequate description of the observed data. One essential method for assessing a model's fit involves plotting data predicted by the model against empirical data (Gelman et al., 1996) to determine whether the model is able to account for key empirical trends. For example, models of the stop-signal task must minimally account for the observed "inhibition function", or the pattern – predicted by the race model – in which individuals' probability of responding on stop trials increases with SSD (Matzke et al., 2018). Models specific to the ABCD task must also account for observed effects of the context independence violation on behavior (choice accuracy decreases at short SSDs, as detailed below).

Second, when exploring the space of possible models that may explain a given cognitive task, the comparison of several quantitative and qualitative features of candidate models is critical (Heathcote et al., 2015; Vandekerckhove et al., 2015). Quantitative comparisons typically involve the use of metrics that quantify tradeoffs between each candidate model's descriptive accuracy and its complexity (Pitt and Myung, 2002). Models of greater complexity (i.e., having greater numbers of parameters or more flexible functional forms) could become "overfit" by capturing unsystematic variance in a specific data set, which hampers their ability to be generalized to broader contexts (Vandekerckhove et al., 2015) and can lead to instability in parameter estimates (Heathcote et al., 2015). Therefore, most quantitative model comparison

metrics, such as the deviance information criterion (DIC) (Spiegelhalter et al., 2002), attempt to prevent overfitting by penalizing for model complexity. However, all quantitative metrics have limitations and must therefore be considered alongside key qualitative features of candidate models (Heathcote et al., 2015). These include the coherence and theoretical precedent of a model's underlying assumptions, its consistency with well-characterized behavioral phenomena, and whether the mechanisms proposed by the model could be eschewed in favor of a more parsimonious explanation (Vandekerckhove et al., 2015).

A third key principle is that a cognitive process model's parameters can only be used to quantify individual or group differences in psychological mechanisms if it is a *measurement model*: one with a sufficiently precise one-to-one mapping between data-generating parameters and parameters estimated by fitting the model to the design from which empirical data will be obtained (Matzke et al., 2020). A cognitive model that accounts for task data well and is supported by the model comparison procedures described above but is unable to provide precise and accurate parameter estimates in a realistic design is unlikely to be useful for measurement purposes in substantive applications. Hence, establishing that a model's parameters can be used to quantify differences in psychological constructs requires parameter-recovery studies, in which data from realistic designs are simulated and the ability of the model and associated estimation procedure to precisely and accurately infer parameter values from simulated data is carefully evaluated (Heathcote et al., 2015).

Several strategies can help researchers improve the estimation and measurement properties of cognitive models. Collecting a large amount of data at the individual level (many separate trials) improves parameter estimation and prevents over-fitting by better matching the complexity of models and data (Smith and Little, 2018). Unfortunately, this is often not possible when developmental and clinical populations of interest can only devote a limited amount of time to a task or when, as in the case of ABCD, the data have already been collected. In these cases Bayesian model estimation methods can be beneficial by providing a principled and quantitative way of using prior knowledge to constrain parameter estimates (Lee and Wagenmakers, 2014). As we will demonstrate, Bayesian methods facilitate the measurement properties of cognitive models for the ABCD stop-signal task, but do not eliminate the need to carefully consider the tradeoffs involved in using empirical data with limited observations at the individual level.

1.2. Cognitive process models of the stop-signal task

As noted above, non-parametric estimation of the independent race model is a foundational measurement tool for providing researchers with an index of inhibitory efficiency in the form of SSRT. However, more recently, several parametric modeling frameworks have been proposed to provide a more comprehensive description of cognitive processes involved in the stop-signal task.

One is BEESTS (Matzke et al., 2013), a Bayesian parametric approach that models the full distributions of finishing times for the go and stop processes as ex-Gaussian distributions (Fig. 1B). The ex-Gaussian distribution, which typically provides an accurate descriptive account of empirical response time (RT) distributions (Heathcote et al., 1991), is the sum of a normal distribution (with mean, μ, and standard deviation, σ) and an exponential distribution (with mean τ), where the latter produces the positive skew characteristic of RTs. Beyond providing a more comprehensive account of empirical data, BEESTS also allows for the estimation of another mechanism that influences inhibitory performance: "trigger failure", in which inattention prevents the stop signal from triggering the inhibitory process, precluding it from entering the race (Fig. 4C). Trigger failures have long been recognized as a possibility (Band et al., 2003; Logan, 1994), but only recently has the BEESTS framework allowed researchers to quantify their non-trivial prevalence (Matzke, Love et al., 2017). This work shows that failing to account for trigger failures can bias non-parametric SSRT and that trigger failures largely explain poor stop-signal task performance in schizophrenia (Matzke et al., 2017) and ADHD (Weigard et al., 2019), which was previously attributed to inhibitory deficits due to bias in non-parametric SSRT estimates. Hence, trigger failure in the BEESTS framework illustrates key advantages of a comprehensive parametric approach for describing the array of factors that contribute to stop-signal task performance and allowing their distinct contributions to effects of interest to be precisely measured.

A second parametric approach draws on evidence-accumulation models (Donkin and Brown, 2018), a widely applied class of models that assume responses on cognitive tasks are the result of gathering noisy evidence for each possible response option over time until a critical threshold of evidence for one option is reached. Within this framework, Logan and colleagues (Logan et al., 2014) proposed an independent racing-diffusion model that assumes stop trial performance results from two go-process accumulators, which gather evidence for each possible response in the go choice task, and a stop accumulator racing towards a threshold (Fig. 4D). This provides a more detailed description of processes that cannot be easily distinguished by ex-Gaussian parameters, including the rate of evidence accumulation for response options versus the amount of evidence an individual requires to make a response (Matzke and Wagenmakers, 2009). Unfortunately, however, more recent work shows that parameter-recovery performance for this model is so poor that it cannot be used to quantify (differences in) the psychological processes represented by the model parameters even when each participant performs thousands of trials (Matzke et al., 2020); despite providing a detailed mechanistic description of performance in the stop-signal task and fitting well to empirical data, the racing-diffusion model cannot be used as a measurement model.

To harness the racing-diffusion model's psychologically plausible description of the go process while preserving the good measurement properties of simpler models, Tanis et al. (2022) recently developed a "hybrid" approach in which the go choice is described with a racing-diffusion evidence-accumulation process but the distribution of finishing times for the stop process is modeled as an ex-Gaussian distribution (Fig. 1E). The resulting racing-diffusion ex-Gaussian (RDEX) model has been shown to provide precise and accurate estimates of parameters that provide an evidence-accumulation account of go choices as well as the parameters that describe the distribution of stopping latencies (including mean SSRT) in realistic task designs (Tanis et al., 2022). Therefore, this framework provides an ideal starting point for developing a measurement model that can account for how the ABCD task design impacts the go process.

1.3. Motivation for the RDEX-ABCD model

In the ABCD stop-signal task, which is completed during functional magnetic resonance imaging (fMRI), children are presented right- or left-facing arrows and instructed to respond with the direction of the arrows via a button box. On stop trials (60 of 360 total trials), an upward-facing arrow, serving as a visual stop signal, replaces the choice stimulus at an SSD determined by a staircase algorithm: increasing or decreasing by 50 ms depending, respectively, on whether or not the last inhibition was successful (Logan, 1994). Replacing the choice stimulus with the stop signal impacts individuals' ability to successfully complete the go choice task by limiting the availability of the choice stimulus to the SSD duration. When SSD = 0, the choice stimulus is never presented and the choice accuracy for responses is therefore necessarily at chance. Choice accuracy increases with SSD to an asymptotic level at SSDs of approximately 0.3–0.4 s (Bissett et al., 2021). In contrast, on go trials the presentation time of the choice stimulus is set to the shorter of 1 s or the participant's response, so the information necessary for accurate responding is equally present on every go trial. This constitutes a violation of the traditional race model's assumption of "context independence", as the go process is clearly not the same between go trials and stop trials of short SSDs, and therefore calls non-parametric SSRT

estimates obtained via the race model into question.

Parametric models such as BEESTS, racing-diffusion, and RDEX usually assume context independence because doing so allows information provided by go trials to improve the accuracy of stop-process parameter estimation. Importantly, however, they do not necessarily need to assume independence if parameters representing how the go process is altered on stop trials can be accurately estimated, as has been recently demonstrated in the context of anticipated response inhibition tasks (Matzke et al., 2021). Therefore, a parametric model that can account for the impact of context independence violations on stop trials with short SSDs could potentially provide valid SSRT estimates in ABCD. Given the difficulties of estimating parameters when data are sparse (in this case, 60 stop trials per person, which is just above the recommended minimum for standard stop-signal tasks (Verbruggen et al., 2019)), the key challenge is to develop a model that both explains the impact of the ABCD task design and allows for accurate recovery of parameters using available ABCD data.

Bissett et al. (2021), who first detailed the context independence violation in ABCD, posit three possibilities for how replacement of the go stimulus with the visual stop signal affects cognitive processing in the ABCD task. The first, "slowed go processing" account holds that the rate of evidence accumulation for the go process is slowed at shorter SSDs because of less efficient information processing. The second, "guessing" account holds that the go process on stop trials is a mixture of guesses and stimulus-driven responses, with 100 % of responses being guesses at SSD= 0 and this proportion gradually decreasing as SSDs get longer. Finally, they posit a "confusion" account, in which participants are unsure of how to respond at short SSDs, leading to a slowing in both the stop and go process as SSDs decrease.

Of these accounts, the first is uniquely well-supported by prior evidence-accumulation modeling of visual masking effects on perceptual choice (Ratcliff and Rouder, 2000; Smith and Ratcliff, 2009; Smith and Sewell, 2013). Such models assume that when briefly presented visual stimuli are interrupted by a mask a noisy representation of the stimulus persists in visual short-term memory after the choice stimulus disappears. The strength and quality of this representation is determined by the stimulus presentation time, which is supported by findings that the average rate of evidence accumulation increases with greater duration (Ratcliff and Rouder, 2000; Smith and Ratcliff, 2009; Smith and Sewell, 2013). Therefore, as the presentation time of the ABCD go stimulus decreases, poorer representations would be expected to cause the rate of evidence accumulation for choices to decrease.

Here, we combine these well-established models of visual-masking effects (Ratcliff and Rouder, 2000; Smith and Ratcliff, 2009; Smith and Sewell, 2013) with the RDEX model (Tanis et al., 2022) to develop "RDEX-ABCD" (Fig. 1E), a modeling framework for describing and measuring the mechanisms through which replacement of the choice stimulus impacts performance on ABCD stop trials. As in the standard RDEX model, we assume that go responses are the result of a race between evidence accumulators that match and mismatch the choice stimulus, rates of which are determined by v_+ and v_- parameters, respectively, and assume that additive infinitesimal Gaussian noise causes evidence totals to fluctuate during accumulation (Tillman et al., 2020). When one of the accumulators reaches a threshold amount, set at parameter B, the corresponding choice is selected. Choice RT is the sum of the time for the winning accumulator to reach threshold and non-decision time (t_0), which is made up of the time to initially encode the stimulus into a form suitable for obtaining decision-relevant evidence and the time to produce a motor response. On stop trials, we adopt the RDEX model's assumption that the go accumulators race a stop process with finishing times that follow an ex-Gaussian distribution defined by parameters μ, σ, and τ. As the ex-Gaussian can take on negative values, we imposed a fixed lower bound of 0.05 s on the stop racer's ex-Gaussian distribution.[1]

We accommodated the violation of context independence in the ABCD task with additional parameters that account for the effect of visual masking on evidence accumulation for go choices at short SSDs. Although estimating go process evidence accumulation rates separately for stop trials of different SSDs could, in principle, address context independence violations in ABCD, such estimates are unlikely to be sufficiently reliable because of the sparsity of stop trials at individual SSDs. Instead, we posit a parsimonious and psychologically plausible growth model in which accumulation rates change with SSD up to an asymptotic level that matches go trials with no masking. That is, we assume context independence only when the effect of masking is negligible. Accumulation rates in the racing-diffusion model can be defined as the sum of two parts, a *discriminative* component, which differentially favors the option that matches the stimulus over the option that mismatches the stimulus, and a *processing speed* component, which favors each option equally and ensures that even very difficult choices with little or no discriminative information produce a timely response (Mazurek et al., 2003; van Ravenzwaaij et al., 2019). We assume that only processing speed (with associated rate v_0) is present when SSD = 0 and that the discriminative component gradually grows with SSD up to an asymptote as the visual short-term memory representation of the stimulus strengthens. The asymptote corresponds to the time at which choice accuracy stops improving with SSD, which is typically in the range 0.3–0.4 s in the ABCD design (Bissett et al., 2021). By assuming that asymptotic discrimination and processing speed are the same for go and stop trials, the model achieves efficient estimation by leveraging information obtained on the more common go trials to constrain parameter estimates for stop trials.

As illustrated in Fig. 1 F, we posit that a simple linear growth process describes the increase in discriminative information at longer presentation durations as changes in matching and mismatching accumulator rates with the same absolute slope g, taking them both from v_0 at SSD = 0 to their respective asymptotic values of v_+ and v_- at longer SSDs. We demonstrate below that, with the relatively few trials per participant available in the ABCD data, more flexible non-linear growth functions that require more parameters to be estimated showed evidence of overfitting, whereas the linear model had both good measurement properties and provided a sufficiently accurate fit.

To summarize, the core RDEX-ABCD model proposes two novel parameters to accommodate the context-independence violation: processing speed (v_0) and the rate of growth in discriminative perceptual information (g) as SSD increases. The remaining parameters are shared with the original RDEX model (Tanis et al., 2022): asymptotic matching (v_+) and mismatching (v_-) rates, non-decision time (t_0) and threshold (B) for the choice accumulators, and the ex-Gaussian parameters for the stop runner (μ, σ, τ). Given prior evidence for the prevalence of trigger failures in stop signal data, we also include a parameter for the probability of trigger failure (p_{tf}) and confirm via model comparison and parameter estimation that this probability is nontrivial in the ABCD sample. Further, we also account for the probability of "omissions", or of failing to respond to the go stimulus, with a "go failure" parameter (p_{gf}). Although relatively rare, go failures have been shown to bias estimates of inhibition in the standard stop-signal paradigm if neglected (Matzke et al., 2019; Tannock et al., 1995).

Of the possibilities Bissett and colleagues (2021) highlight, the RDEX-ABCD model explicitly formalizes their slowed-go-processing account. RDEX-ABCD also features a version of guessing in that evidence accumulation rates for matching and mismatching choices are identical on 0-SSD trials, leading to guess responses that are driven only by processing speed. In contrast, however, Bissett and colleagues'

[1] Assuming bounds from 0 to 0.1 s did not impact results as the probability of such fast stop processes was negligible.

Table 1
Prior probability distributions used for Bayesian estimation of the RDEX-ABCD model's parameters. All priors were independent truncated normal distributions with the bounds, location parameters and scale parameters listed in the table. Location and scale parameters for the "broad" prior distributions were selected a priori. Location and scale parameters for the "informative" prior distributions were derived from a hierarchical Bayesian version of the model that was fit to a subset of 300 ABCD participants using procedures described in the text. Note that priors for the p_{gf} and p_{tf} parameters are represented on the probit scale.

	Bounds		*Broad Prior Distribution*		*Informative Prior Distribution*	
Parameter	lower	upper	location	scale	location	scale
t_0	0.100	1.000	0.150	0.100	0.027	0.068
B	0.000	∞	1.000	1.000	1.444	0.308
v_+	0.000	∞	3.000	2.000	3.334	0.550
v_-	0.000	∞	1.000	2.000	-0.045	0.525
v_0	0.000	∞	2.000	2.000	2.732	0.634
g	-∞	∞	3.000	2.000	2.958	0.862
μ	0.000	2.000	0.500	0.300	0.226	0.036
σ	0.000	0.500	0.050	0.100	0.012	0.095
τ	0.000	0.500	0.100	0.100	0.014	0.032
p_{gf}	-∞	∞	0.000	1.000	-1.944	0.517
p_{tf}	-∞	∞	0.000	1.000	-1.892	0.764

original guessing-based account assumes guesses are mixed with fully stimulus-driven responses even at short SSDs. We did not adopt this approach given findings that little discriminative information is available at short presentation times (e.g., 0.05–0.10 s (Smith and Ratcliff, 2009)). Although we initially considered extensions to RDEX-ABCD that could allow SSRT to differ at shorter SSDs, as suggested by the proposed "confusion" account (Bissett et al., 2021), we provide evidence that the core RDEX-ABCD model parsimoniously explains key trends in the ABCD data without the need for such complex extensions.

The next sections describe the application of RDEX-ABCD and comparison models to empirical ABCD data. We show that the model provides a good description of key trends on stop trials, including the SSD effect on choice accuracy that is the hallmark of the context independence violation, and shows clear advantages relative to comparison models. We next use simulations of the RDEX-ABCD model to demonstrate that ignoring growth in perceptual choice information in the ABCD design can lead to consequential biases in non-parametric SSRT estimates. Crucially, we show that our model can overcome these measurement difficulties and allow data from the ABCD sample to be used to obtain unbiased estimates not only of SSRT, but also other processes of interest.

2. Materials and methods

2.1. Participants, inclusion criteria, and subsample selection

The ABCD Study is a multi-site collaboration that has recruited a diverse sample of 11,875 children between the ages of 9 and 10 from 21 study sites across the United States. Details of the general study design, recruitment procedure, and fMRI protocol (including the stop-signal task design) are available elsewhere (Casey et al., 2018; Garavan et al., 2018). All child participants in ABCD provided verbal assent while parents provided written informed consent. The ABCD data repository, which is openly available via the National Institute of Mental Health Data Archive (NDA), grows and changes over time. As the combination of Bayesian estimation procedures with model comparison analyses is computationally intensive, fitting all models to the full sample during model development and testing would have been unnecessarily time intensive. Therefore, we randomly selected a subset of 900 individuals (300 for the derivation of informed priors and 600 for the primary analyses) whose baseline session stop-signal data met our performance-based inclusion criteria. We selected this initial subset from the set of individuals whose baseline stop-signal data were included in the ABCD Annual Release 1.1 dataset (n = 4521, NDA Study 576, DOI 10.15154/1412097, available at: https://ndar.nih.gov/study.html?id=576).

Of the 4521 individuals in Release 1.1, we found that 3436 had complete behavioral data available from the stop-signal task that met basic validity checks (two imaging runs, 360 recorded trials, and participant responses detected for one or more go trials). Performance exclusion criteria applied to ensure adequate engagement with the task and consistency with race model assumptions (detailed in Supplemental materials) led to the exclusion of 16.8% of the initial sample of 3436, leaving 2859 participants available for modeling analyses. Out of this remaining sample, we sought to randomly select a subsample of non-sibling participants from across multiple data collection sites. We first determined, for each data collection site, how many individual family identification numbers (family IDs) had stop-signal data available for modeling from at least one child. We then selected the 6 data collection sites with the greatest number of available family IDs and randomly sampled (without replacement) 150 family IDs from each of these 6 sites. We included single children from each family for which only one child was available and randomly sampled a sibling from each family for which multiple children's data were available. This led to a subsample of 900 participants from independent families (150 per site) that was further split into a 300 participant (50 per site) subsample included in hierarchical model fits to estimate parameter priors and a 600 participant (100 per site) subsample included in the main analyses. Supplemental Table 1 displays basic demographic characteristics of the 900 participants selected for analysis.

2.2. Models evaluated

The parameterization of the "core" 11-parameter RDEX-ABCD model is described in the text above. We also estimated several comparison models to explore the model space for possible improvements on RDEX-ABCD and to contrast it with models of lesser complexity (those that do not attempt to account for context independence violations) and greater complexity (e.g., those positing non-linear effects of SSD on the go process).

First, to gauge the importance of accounting for the ABCD task's context independence violation in the model, we estimated a "go-independence" comparison model that had an identical structure to RDEX-ABCD except for assuming that the go process across stop trials of all SSDs had accumulation rates identical to the accumulation rates on go trials (i.e., v_0 and g were not estimated, leading to a simpler 9-parameter model). Second, we explored the possibility that a more flexible, non-linear growth function would be better than RDEX-ABCD's simple linear growth model at describing the process by which drift rates grow from v_0 at SSD= 0 to their asymptotic values at longer SSDs. Specifically,

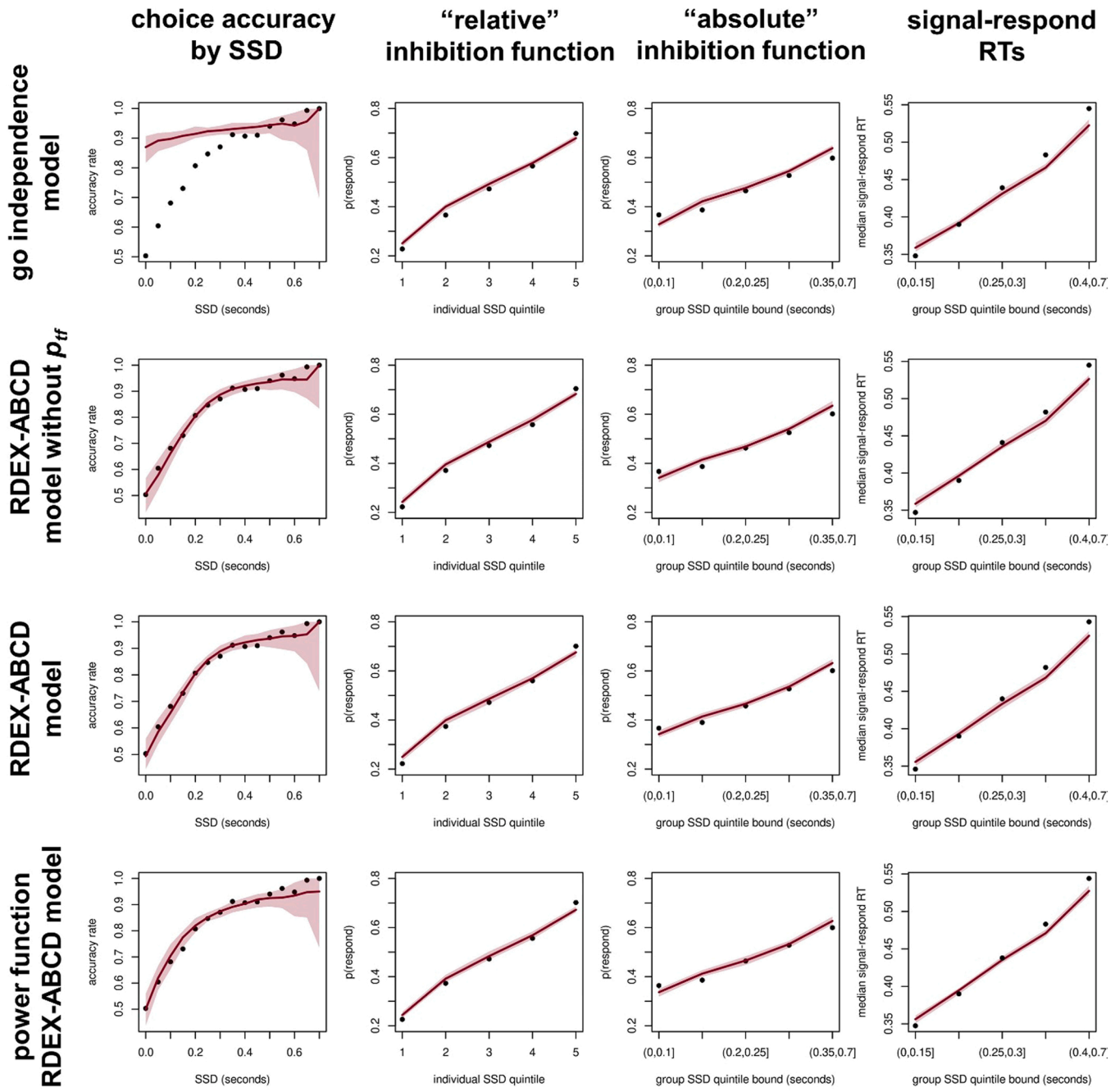


Fig. 2. Posterior predictive plots showing the RDEX-ABCD model and comparison models' median predictions (red line) and 99% credible intervals (CIs) of predictions (red shading) for key trends in the ABCD stop-signal task data, overlayed with empirical values (dots). All models were fit with informative priors. Choice accuracy for signal-respond trials is binned by SSD, showing the hallmark effect of the ABCD task's violation of context independence: the systematic increase in choice accuracy with go stimulus presentation times. For the "relative" inhibition functions, SSDs are binned at the individual level to account for individual differences in performance. For the "absolute" inhibition functions and plots of median signal-respond response time (RT) by SSD, SSD bins are created by collapsing SSDs for observed trials across the whole group.

we estimated a model in which growth was governed by a generalized power function, $(x+k)^a$, with k and a being free parameters, leading to a 12-parameter model overall (other models, including exponential forms, were explored but did not produce differing results). Finally, to gauge the importance of accounting for trigger failure in the ABCD data, we alternately estimated the RDEX-ABCD model, as well as the two comparison models detailed above, with and without the p_{tf} parameter.

2.3. Model estimation

We implemented all models within Dynamic Models of Choice (DMC), a free set of R functions for Bayesian estimation and simulation of evidence accumulation and stop-signal task models (Heathcote et al., 2019). Following previous work (Matzke et al., 2017; Weigard et al., 2019), parameters for the probability of trigger failure (p_{tf}) and "go" failure (p_{gf}) were projected onto the real line during model estimation using a probit transformation and later transformed back to the natural scale for interpretability. Before all modeling analyses, trials with RTs < 0.15 s were excluded as fast guesses and trials with RTs > 1.5 s were excluded as abnormally slow responses (these exclusions removed $<1\%$ of trials).

Each model was fit twice, using individual-level Bayesian estimation,

Table 2
Model comparison metrics for all models evaluated under both broad and informative priors, alternately fit with and without the trigger failure probability (p_{tf}) parameter. All model comparison metrics are zero-based, by subtracting the lowest value from all other values in each comparison, to improve clarity. *np* = number of parameters in each model; DIC = deviance information criterion; BPIC = Bayesian predictive information criterion.

		Broad Prior Fits		*Informative Prior fits*	
Model	*np*	DIC	BPIC	DIC	BPIC
go-independence, no p_{tf}	8	2667	2000	4291	4332
go-independence with p_{tf}	9	3094	2965	3200	3109
linear RDEX-ABCD, no p_{tf}	10	0	0	1016	1260
linear RDEX-ABCD with p_{tf}	11	312	797	315	603
power function RDEX-ABCD, no p_{tf}	11	185	128	1075	1216
power function RDEX-ABCD with p_{tf}	12	494	913	0	0

to individuals in the main 600-participant subsample: once with broad, relatively uninformative priors and a second time with informative priors derived from hierarchical model fits to the independent subsample of 300 participants. The latter procedure was used to prevent over-fitting by providing additional constraints on individuals' parameter estimates. Although the use of a single hierarchical model for the entire sample could accomplish the same goal of constraining individual-level parameter estimates, such a procedure has two downsides in the present context. First, implementation of fully hierarchical models within the full ABCD sample would be computationally intensive, likely exceeding reasonable time and memory limits when conducted on most research computing systems. Second, individual-level parameters from hierarchical models are inappropriate for inclusion in many follow-up statistical analyses (Boehm et al., 2018). Therefore, the two-stage procedure used here helps prevent over-fitting without these drawbacks.

All priors were truncated normal distributions. Locations, scales, and bounds of the broad and informative priors for the RDEX-ABCD model are displayed below in Table 1; the corresponding values for all other models evaluated are reported in Supplemental Table 2-6. Hierarchical models fit to the 300-person independent subsample that was used to obtain informative priors treated individual-level parameters as random effects described by group-level truncated normal distributions that are defined by location and scale hyperparameters. Priors for the location hyperparameters were the same as the broad priors used for the initial individual-level fits (Table 1; Supplemental Table 2–6) while priors for all scale hyperparameters consisted of exponential distributions with a scale of 1. Following the estimation of posterior distributions from the hierarchical model for the 300-person subsample, we collapsed the individual-level posterior samples across all 300 individuals into a single vector for each model parameter. We then fit a truncated normal distribution to samples from each parameter vector using maximum likelihood estimation. The resulting location and scale parameters of the fitted truncated normal distributions were used to specify informative priors for the remaining 600 participants.

All hierarchical and individual-level Bayesian parameter estimation procedures used the differential-evolution Markov chain Monte Carlo (DE-MCMC) algorithm to sample from the posterior, which is suitable for evidence-accumulation models and other models that tend to have correlated parameters (Turner et al., 2013). Sampling used a number of chains that, by DMC default, was three times the parameters in a given model (e.g., 33 chains for the 11-parameter RDEX-ABCD model). Each sampling run featured an initial burn-in period that included a migration (Turner et al., 2013) step (with 5 % probability in individual-level fits and 2.5 % probability in hierarchical fits) and lasted until no chains were repeatedly "stuck" in low likelihood locations, as determined by an automated function in DMC with default settings. Next, a second burn-in period was started with migration turned off, lasting until chains for all parameters had converged to the posterior mode. Final convergence was defined as values of < 1.10 for the Gelman-Rubin diagnostic statistic (Gelman et al., 1992), and was corroborated via visual inspection of the chains (Lee and Wagenmakers, 2014). Following initial sampling, posterior samples for individuals' mean SSRT were computed by simulating 10,000 finishing times from ex-Gaussian distributions defined by all sets of μ, σ, and τ posterior samples and then, for each set of samples, taking the mean of the simulated finishing times greater than the .05 s lower bound. These SSRT samples were then aggregated to form SSRT posterior distributions.

2.4. Assessment of model fit and model comparisons

Posterior predictions were generated by drawing 100 samples from the joint posterior for each model and simulating data predicted by the model for each participant and each of the 100 sets of parameter samples. The predicted data for each of the 100 sample sets were averaged across participants within each set to obtain summary values of interest (e.g., average accuracy rates and probability of inhibition at specific SSDs). The medians and 99 % credible intervals (CIs) of these group-average predictions were then estimated with 100 draws of posterior samples and compared to the empirical group-average values in posterior predictive plots (Gelman et al., 1996).

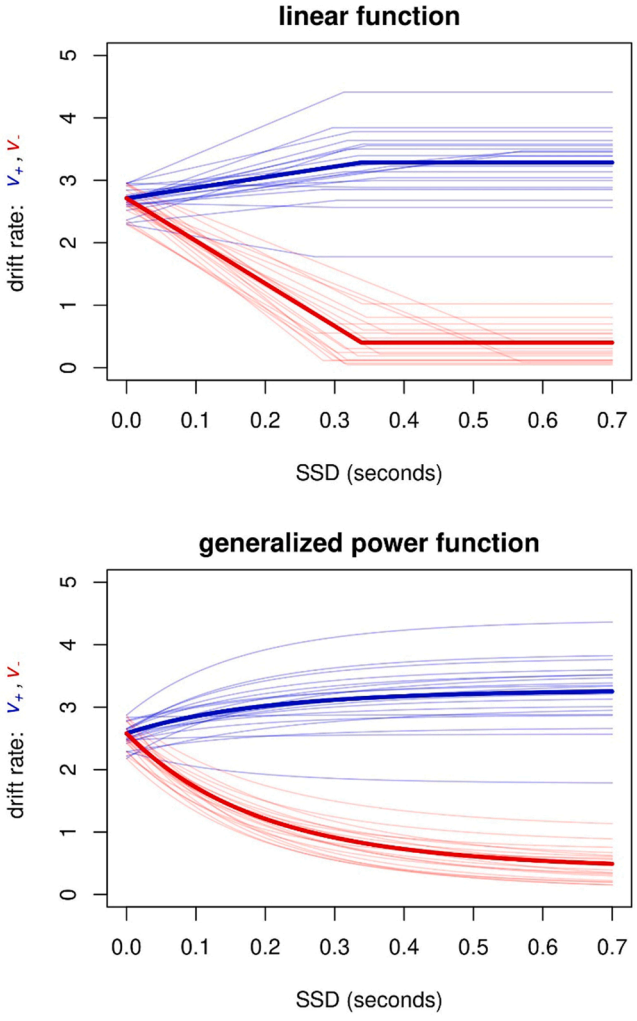


Fig. 3. Empirical growth patterns of matching (blue lines increasing from SSD = 0) and mismatching (red lines decreasing from SSD = 0) "go" process accumulator rates by stop-signal delay (SSD) for the sample average parameter estimates (thick lines) and for parameter estimates from 20 randomly drawn participants (thin lines) to illustrate individual variability.

Table 3
Average RDEX-ABCD model parameter estimates and model-based mean stop signal reaction time (SSRT) estimates for the 600-participant group and the 99% posterior credible intervals for these group averages. The tight credible intervals for all parameters reflect the large (600 person) sample size on which the group average is based.

Parameter name	Parameter definition	Group average	Credible interval	
			0.50 %	*99.50 %*
t_0	Go non-decision time	0.137	*0.134*	*0.138*
B	Go evidence threshold	1.417	*1.407*	*1.427*
v_+	Go match rate	3.285	*3.270*	*3.308*
v_-	Go mismatch rate	0.401	*0.383*	*0.441*
v_0	Go processing speed	2.713	*2.662*	*2.767*
g	Perceptual growth rate	2.953	*2.883*	*3.060*
μ	Stop ex- Gaussian normal mean	0.229	*0.225*	*0.232*
σ	Stop ex-Gaussian normal SD	0.076	*0.071*	*0.090*
τ	Stop ex-Gaussian exponential mean	0.030	*0.028*	*0.034*
p_{gf}	Probability of go failure	0.027	*0.026*	*0.028*
p_{tf}	Probability of trigger failure	0.037	*0.031*	*0.041*
SSRT	Mean stop signal reaction time (computed by simulation)	0.268	*0.265*	*0.276*

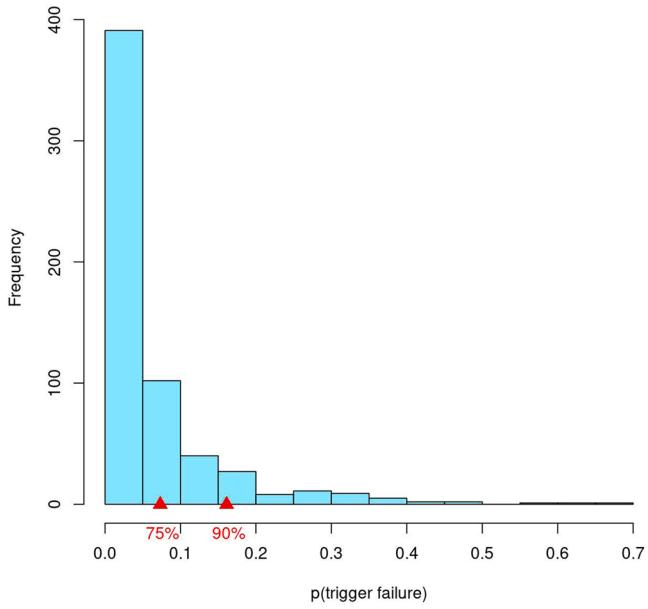


Fig. 4. Histogram of individual-level estimates (posterior medians) for the probability of trigger failure in the 600-person ABCD subsample. Red triangles denote the 75th and 90th percentiles.

In addition to assessing models' descriptive accuracy with posterior predictive plots, we applied two model comparison metrics that reward models for goodness-of-fit but also penalize models for greater complexity: the deviance information criterion (DIC) (Spiegelhalter et al., 2002) and Bayesian predictive information criterion (BPIC) (Ando, 2011).

3. Results

In this section, we: (1) assess the ability of RDEX-ABCD and a set of comparison models to describe empirical stop-signal task data from ABCD; (2) outline RDEX-ABCD's account of the context independence violation and substantive interpretations of its parameters; (3) provide examples illustrating how variation related to the violation can confound inferences about inhibitory ability if unaccounted for; and (4) demonstrate that RDEX-ABCD can be effectively used to measure SSRT and other parameters of interest from ABCD data.

3.1. Model fit to ABCD data

Posterior predictive plots in Fig. 2 detail models' ability to account for the increase in empirical choice accuracy with SSD, the hallmark of the context independence violation on the ABCD task, as well as two other empirical trends that are conventionally used as benchmarks of goodness-of-fit in the stop-signal modeling literature (Matzke et al., 2018).

The first trend relates to the probability of responding as a function of SSD, the "inhibition function", which the race model predicts to be increasing. Following prior work (Heathcote et al., 2019; Weigard et al., 2019), we plotted the average inhibition function using relative SSD bins based on quintiles (i.e., five equal-probability groups of ordered SSDs) for each individual participant in order to account for individual variation in inhibitory performance. Forming SSD bins based on absolute times confounds within- and between-participant performance (as different individuals have different ranges of SSDs due to the adaptive tracking algorithm), flattening the average inhibition function, although we also display plots that use this absolute binning procedure for comparison.

The second trend is that "signal-respond" RTs (RTs for responses on stop trials) increase with SSD because the go RT distribution is censored by successful stopping to a lesser degree at longer SSDs (Colonius et al., 2001; Matzke et al., 2018). In contrast to the inhibition function, this trend is best represented by binning RTs according to their absolute values rather than by individualized/relative bins. Because the task's staircase algorithm leads individuals of different ability levels to be presented with very different sets of SSDs, individualizing the bins causes within-bin variability in SSD to be large compared to between-bin variability in SSD. The resulting increase in within-bin variability in SSD, which in turn increases variance in the censoring of RT within the bins, leads plots of signal-respond RT by relative SSD bins to be mostly flat and uninformative. Therefore, following prior work (Heathcote et al., 2019; Weigard et al., 2019), we plotted the average trend using absolute quintile bins calculated from SSDs collapsed across signal-respond trials of all individuals.

All models that attempted to account for the impact of context independence violations on the go process provided excellent descriptions of the pattern in which choice accuracy rates increase with SSD. Differences between predictions of the model that assumed growth was described by a two-parameter power function and models assuming a simpler linear function were relatively subtle. In contrast, the go-independence model displayed increasingly gross misfits to empirical accuracy rates as SSD decreased from .35 s. Therefore, RDEX-ABCD's assumption of growth in go process perceptual evidence quality appears sufficient to provide a compelling account of the behavioral hallmark of context independence violations in ABCD.

Posterior predictive plots of the inhibition function indicated that all models also provided a good account of response probabilities and their pattern of increasing as SSDs grow longer.[2] All models similarly displayed generally good fit to SSD-related increases in median signal-respond RT, although they slightly underpredicted the slope of the increase and the absolute value of the highest SSD quintile. These misfits may indicate unexplained or contaminant processes at very long SSDs. However, as the absolute level of misfit is relatively small (about .025 s

[2] We note that when Bissett et al. (2021) plotted inhibition functions while using each specific SSD to form "absolute" bins they found an apparent increase in responding on 0 s SSD trials relative to other short SSD trials, which is unexpected given race model assumptions. However, we show in Supplemental materials that this uptick in responding is likely an artifact of averaging across individuals who have different levels of inhibitory performance and that the RDEX-ABCD model can account for this artifactual uptick when it is unconstrained by priors.

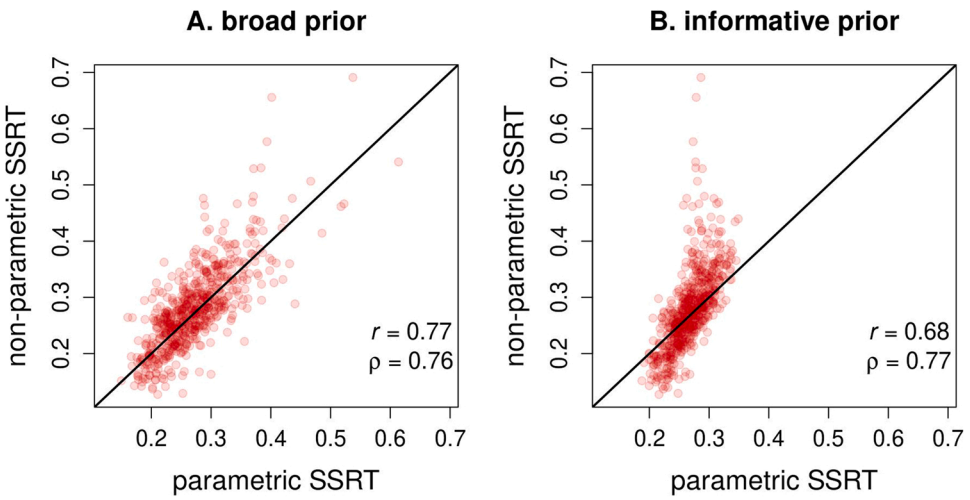


Fig. 5. Comparison of non-parametric SSRT estimates with parametric estimates obtained from RDEX-ABCD, both when the model is estimated with broad, uninformative priors (**A**) and when it is estimated with narrower priors informed by a hierarchical model fit (**B**). The black line represents where dots would fall along if the relation between the two sets of estimates was perfect. Correlation coefficients for each relation, including both Pearson's *r* and Spearman's ρ, are displayed in the bottom right corner of each plot.

at most), and as the general increasing trend is well-described, all models appear to provide an adequate account of signal-respond RTs.

Despite the go-independence model clearly providing an inadequate account of go choice accuracy on stop trials with short SSDs, its description of the inhibition function and signal-respond RT is notably similar to those of the alternate models that explicitly accounted for context independence violations. This suggests that, although changes in evidence quality at shorter SSDs caused by the violation have a substantial impact on choice accuracy, their impact on the average latency of go process RTs is less pronounced. As go-process RTs would be expected to impact individuals' probability of responding and signal-respond RTs on stop trials regardless of the accuracy of the go responses, this explanation could account for the apparently subtle impact of the context independence violation on these group-level trends.

Quantitative model comparison metrics (DIC and BPIC), summed across participants and zero-based for clarity (by subtracting the lowest DIC/BPIC value from all other values in each comparison), are reported in Table 2 for all models fit to the ABCD data with both broad and uninformative priors. Unsurprisingly, these metrics indicate that the go-independence model displays unambiguously poorer fit than all other models, consistent with its clear difficulty accounting for choice accuracy at short SSDs (Fig. 2). These metrics' support for the inclusion of trigger failure and nonlinear patterns of perceptual growth across SSDs was more ambiguous. Inclusion of the p_{tf} parameter led to better DIC/BPIC values when models were fit with informative priors but poorer DIC/BPIC values when models were fit with broad priors. Similarly, the RDEX-ABCD model with a simple linear growth function displayed better DIC/BPIC values than the power function model when broad priors were used, but the power function model was more frequently supported when informative priors were used (except in considering DIC when p_{tf} is absent from both models). In both cases, this ambiguity is consistent with the posterior predictive plots in Fig. 2, which indicate that inclusion of p_{tf} and nonlinear growth parameters in the RDEX-ABCD framework does not lead to visibly unambiguous improvements in model fit.

In sum, evaluations of model fit suggest that the RDEX-ABCD framework's assumption of perceptual growth in go process evidence quality provides an excellent account of the pattern of increasing choice accuracy rates by SSD, the hallmark of context independence violations in the ABCD data. Therefore, in the section below, we go on to interpret parameters of the RDEX-ABCD model. As evidence for the inclusion of nonlinear perceptual growth patterns and trigger failures was more ambiguous, we also consider the practical consequences of including and interpreting these mechanisms in the model.

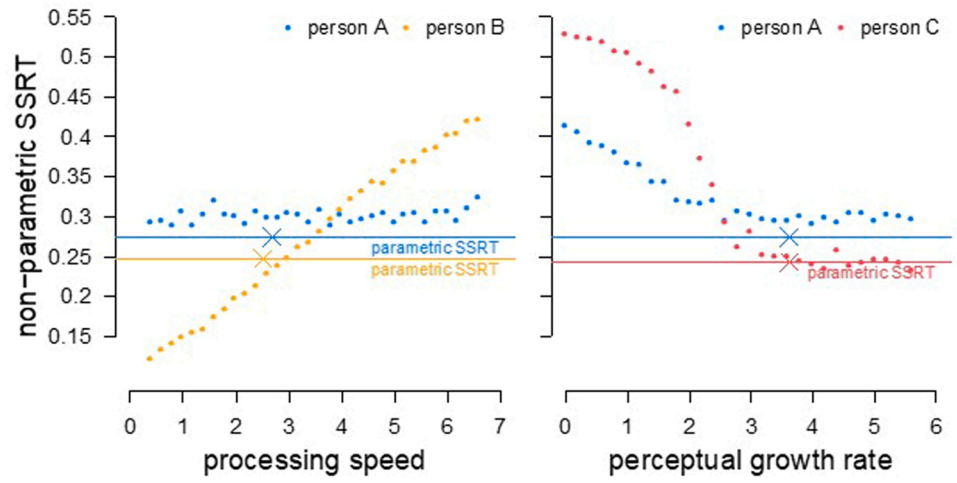


Fig. 6. A demonstration of the influence of processing speed and perceptual growth rate on the non-parametric SSRT of two individuals based on 15,000 simulated "go" and 5000 simulated "stop" trials each. Persons A and B (left) have a similar level of speed as indicated by the crosses, but a differing SSRT. When speed is varied, the estimated non-parametric SSRT is affected even though this change should not lead to a different SSRT. Similarly, persons A and C (right) have a similar level of perceptual growth rate, and varying this rate also affects the estimated non-parametric SSRT. Both cases result in two regions leading to qualitatively different conclusions. In reality, person A has a higher SSRT than persons B and C, and this true parametric SSRT (horizontal lines) is independent of both speed and perceptual growth rate. Note that with the large number of trials in these simulations RDEX-ABCD parametric SSRTs are essentially perfect, and so the horizontal lines correspond to the true data generating values.

3.2. Practical interpretation of the RDEX-ABCD model

Fig. 3 shows how SSD in the ABCD stop trial design affects choice processing rates on average, and for 20 randomly selected individuals, under the alternate assumptions of growth governed by linear versus generalized power functions. In the linear model, matching (v_+) and mismatching (v_-) accumulator rates grow from being (necessarily) identical at 0 s SSDs to being equivalent to their go trial levels at SSDs between 0.3 s and 0.4 s for most participants, consistent with the SSD region where accuracy becomes asymptotic (Bissett et al., 2021). The generalized power function, although agreeing with the linear model's description of 0 s SSD and asymptotic drift rates, posits that change in drift rate is most rapid at the shortest SSDs and gradually decelerates as SSDs grow longer.

When combined with the indications, detailed in the section above, that the generalized power function does not clearly improve model fit relative to the simpler linear function, there are two reasons why estimation of the more complex model may not be warranted in most applications to ABCD data. First, the generalized power function parameters are not as readily interpretable as the linear slope parameter *g*, which provides straightforward estimates of individuals' rate of change in go process evidence quality across SSD durations. Second, we found that estimates of the other model parameters were nearly perfectly correlated between the generalized power and linear growth model versions (Supplemental materials), indicating that the specific growth function assumed has only a trivial impact on substantive inferences about the go and stop processes. Therefore, unless the precise shape of perceptual growth is of interest for a given research question, the core RDEX-ABCD model assuming linear growth appears best suited for applications in ABCD.

Table 3 displays group average parameter estimates for this model. The 6 go parameters indicate that go failures are rare (2.7%), and that the mean times for matching and mismatching go accumulators to finish ($t_0 + B/v_+$ and $t_0 + B/v_-$, respectively) are 0.57 s and 3.67 s, which would cause observed correct RTs to have a mean of 0.55 s and observed error RTs to have a mean of 0.58 s. In comparison, when there is no discriminative information (v_0 only), the mean finishing time for the go race is slightly faster, leading observed RTs (which would be split evenly between correct and error responses) to have a mean of 0.52 s. The average SSRT estimated by the model is 0.268 s, considerably slower than most previous BEESTS estimates foradult participants (Matzke et al., 2017; Skippen et al., 2019), but comparable to estimates from children of similar age (0.243 s for healthy 8–12 year-olds) (Weigard et al., 2019).

The 3.7 % trigger failure rate is on the low end of previous findings with non-clinical groups (Matzke et al., 2017; Weigard et al., 2019), suggesting participants were generally attentive to the task. This relatively low rate, on average, is consistent with the ambiguous evidence from group-level model comparisons for the inclusion of the p_{tf} parameter in the model. However, inspection of individual-level parameter estimates suggests that trigger failures play a substantial role in the performance of a subset of participants (Fig. 4), with the top quartile of individuals displaying rates higher than 7.3% and the top decile displaying rates higher than 16.1 %. Coupled with prior evidence that trigger failures are important for explaining individual differences in clinical phenotypes (Matzke et al., 2017; Weigard et al., 2019) and can bias SSRT estimates if unaccounted for (Matzke et al., 2017), the presence of individuals with non-trivial trigger failure rates indicates that estimation of p_{tf} in the ABCD data set is warranted. Indeed, linking these salient individual differences to neural and clinical covariates would likely be a productive area of future work that could shed light on the contributions of attentional processes to inhibitory performance.

In sum, RDEX-ABCD not only describes ABCD data well by adding a parsimonious explanation for the impact of context independence

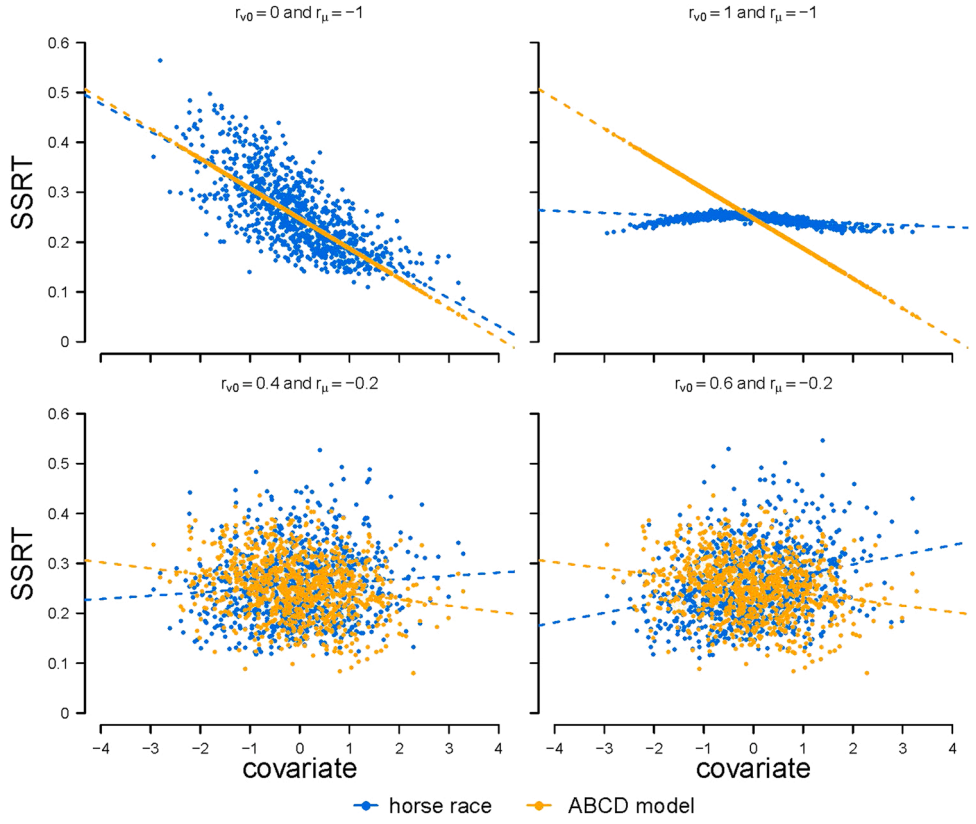


Fig. 7. Demonstrations of covariate confounds influencing non-parametric SSRT estimates in a simulated correlational analysis. A covariate was created that has a relation with both processing speed (v_0) and SSRT (through the μ parameter) for 900 simulated participants. The scatterplots display the observed relations between SSRT and the covariate when non-parametric methods (blue) versus the parametric RDEX-ABCD model (orange) are alternately used to estimate SSRT.

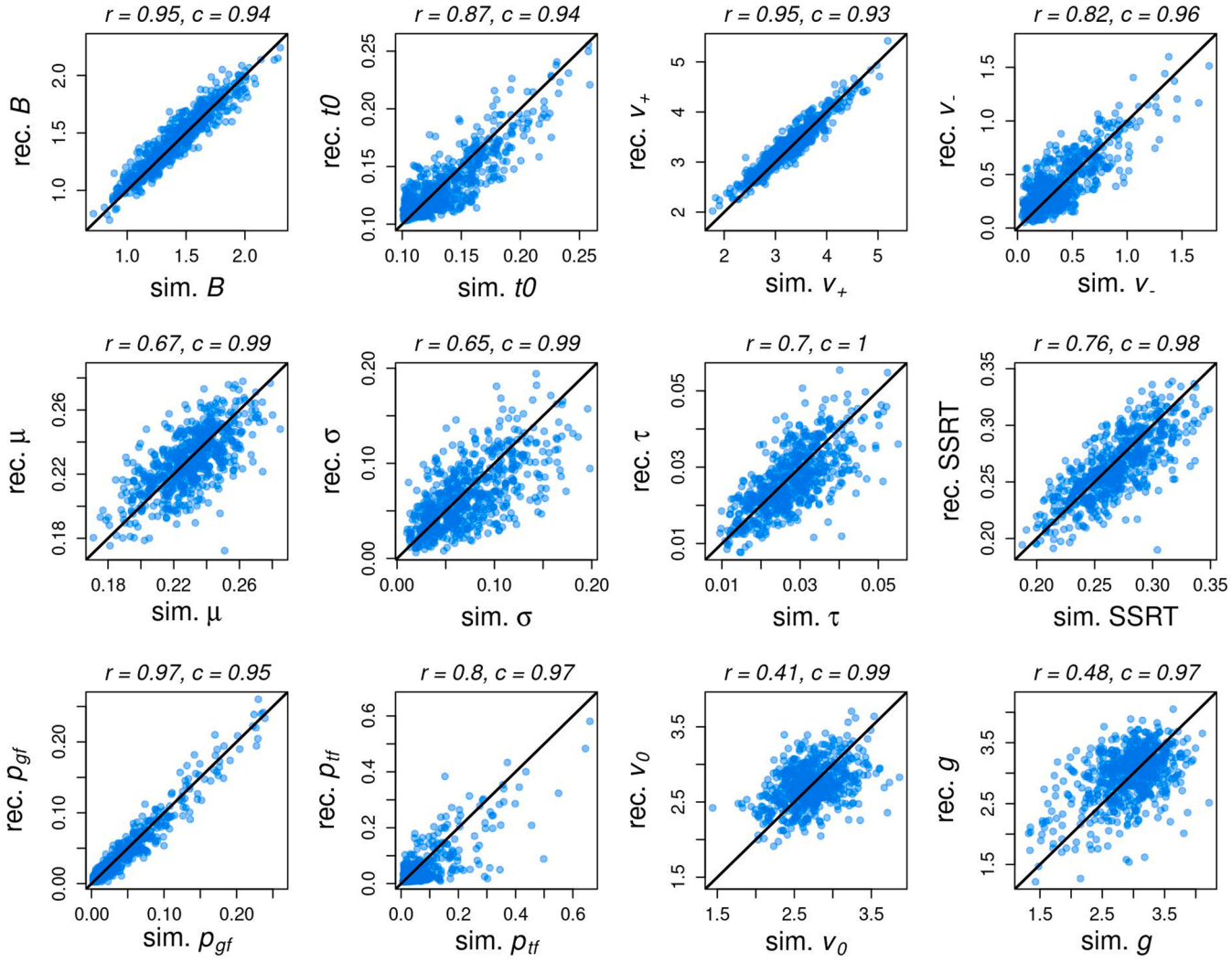


Fig. 8. Results from the parameter recovery study in which informed priors were used to estimate parameters (both the initial parameter values estimated from empirical data and the parameter values recovered from simulated data). Scatterplots illustrate the relations between the simulated ("sim.") and recovered ("rec.") parameter values as compared to the diagonal solid line indicating perfect recovery. Numbers above each plot report the correlation coefficient (*r*) for each relation and the posterior coverage proportions (*c*) for each parameter, which indicate the proportion of data-generating parameter values that fall within the 95 % posterior credible interval for the parameters recovered from the generated data.

violations to the existing RDEX model framework (Tanis et al., 2022), but also yields interpretable parameters that represent components of the stop process (including effects of trigger failure), the go process, and the impact of independence violations on the latter. We next use the model to quantify the degree of bias that could occur in non-parametric SSRT estimates that are based on the original race model, which does not account for as many features of the stop process or for context independence violations.

3.3. Quantifying biases in non-parametric SSRT

We first assessed the relationship between the parametric SSRT estimates from RDEX-ABCD and those from the "best practice" non-parametric integration method (Verbruggen et al., 2013) that also accounts for go trial omissions (Verbruggen et al., 2019). As shown in Fig. 5, the two measures shared around 60% of their variance (as determined by squaring Pearson's *r*) and had a close to one-to-one relationship when the ABCD model was estimated with broad priors. This shared variance was reduced, and the one-to-one relationship disappeared, when priors informed by the hierarchical model fit were used. The latter result reflects hierarchical "shrinkage" (Efron and Morris, 1977; Gelman et al., 2013), which produces better estimates on average by pulling poorly constrained outlying estimates closer to the group mean, as is evident in the right panel of Fig. 5. Regardless, the rank ordering of participants' parametric SSRTs is relatively well-preserved in the non-parametric estimates. In fact, the rank correlation (Spearman's ρ) between non-parametric and parametric estimates was nearly identical when broad and informative priors were alternately used to fit the model.

Although the ordinal correspondence between RDEX-ABCD and non-parametric SSRT estimates suggests that the latter can be used to study individual differences in inhibition, confounding factors that are not accounted for in non-parametric methods may nonetheless lead to misleading inferences. Indeed, we found in simulation studies (methods described in Supplemental) that biases in non-parametric estimates can be consequential for inferences at both the individual and group levels. We present an example at each level to demonstrate that differences in processing speed (v_0) and perceptual growth rate (g) can be mistaken for differences in SSRT, as estimated by the non-parametric method. Note that the parameter values used in these simulations, and ranges over which we vary parameters, are representative of those found in the ABCD data. We also point out that trigger failures are already known to

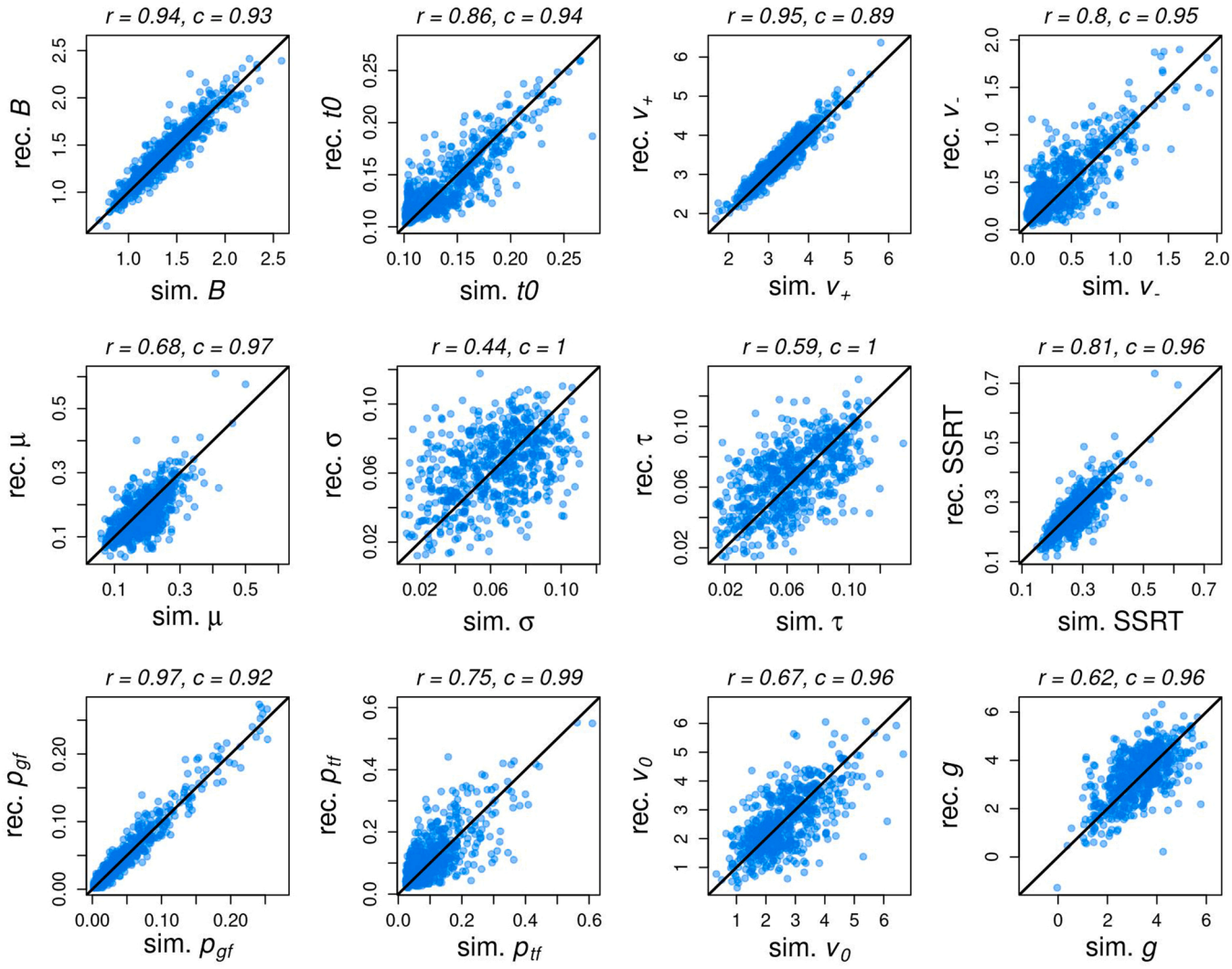


Fig. 9. Results from the parameter recovery study in which broad priors were used to estimate parameters (both the initial parameter values estimated from empirical data and the parameter values recovered from simulated data). Scatterplots illustrate the relations between the simulated ("sim.") and recovered ("rec.") parameter values as compared to the solid diagonal line indicating perfect recovery. Numbers above each plot report the correlation coefficient (*r*) for each relation and the posterior coverage proportions (*c*) for each parameter, which indicate the proportion of data-generating parameter values that fall within the 95 % posterior credible interval for the parameters recovered from the generated data.

bias non-parametric SSRT (Matzke et al., 2017), and so the trigger failure parameter can partially account for differences between parametric RDEX-ABCD and non-parametric SSRT estimates. However, simulation studies reported in this section hold this parameter constant.

The first example shows that ignoring either speed or growth-rate differences can lead to reversed SSRT effects when making pairwise comparisons among three individuals. Person A has a higher true parametric SSRT than the others, but roughly the same speed as B and a similar perceptual growth rate to C. Fig. 6 shows the results, where horizontal lines indicate the true parametric SSRT for each person, crosses correspond to the true speed (left panel) and perceptual growth rate (right), and dots show non-parametric SSRT estimates for each simulated dataset. Varying speed had hardly any effect on the non-parametric SSRT estimates of A, whereas a positive linear relation was found for B, creating two qualitatively different regions. When speed was below approximately 4, A had a higher non-parametric SSRT than B, whereas this order was reversed when speed exceeded four. By not taking processing speed into account, the non-parametric SSRTs would result in incorrect conclusions in the latter region. Similarly, varying perceptual growth rate had different effects on A and C, and again created two qualitatively different regions below and above around 2.5. Non-parametric SSRT would lead to incorrect conclusions in the lower region.

Both cases demonstrate that overlooking individual differences in processing speed and perceptual growth rate can lead to incorrect inferences about the relative inhibitory abilities of two individuals. Unless individual differences in these factors are considered by using the RDEX-ABCD model, there is potential for the non-parametric method to identify putative differences in inhibitory ability where none exist, and even for one individual to be identified as better at inhibition than another when in fact the opposite is true.

The stop-signal task is often used to identify differences in SSRT between groups or experimental conditions, or to investigate whether a covariate (e.g., ADHD symptoms or activity in a particular brain region) is related to inhibitory ability. Our second simulation shows that a covariate related to both speed (v_0) and SSRT (through the μ parameter) may cause non-parametric SSRT to provide reversed conclusions in a realistic analysis. Fig. 7 shows an illustrative selection of results for 900 simulated participants. The top row shows two extreme situations in which SSRT was perfectly negatively correlated with the covariate, but there was alternately no (left plot) or a perfect positive (right plot) correlation between the covariate and processing speed. Two situations

with realistic intermediate covariate correlations are presented in the bottom row.

Both non-parametric and parametric SSRT correctly showed strong negative correlations with the covariate when only SSRT was perfectly related to the covariate (top-left). However, their results clearly diverged when speed also had a perfect positive correlation with the covariate (top-right). In this case, parametric SSRT had the expected strong negative relation with the covariate, while non-parametric SSRT estimates hardly varied over different levels of the covariate. Non-parametric SSRT estimates failed to pick up the negative relation between the covariate and true SSRT because of the counter-balancing effect of the positive speed correlation on non-parametric SSRT.

The bottom row of Fig. 5 demonstrates that intermediate covariate correlations, closer to those expected in empirical data, can also result in reversed effects. In both scenarios, RDEX-ABCD accurately detected a small *negative* relation between the covariate and SSRT. From the non-parametric estimates, however, we would conclude that there is a *positive* relation between the covariate and SSRT because the positive correlation with processing speed reverses this trend.

These results provide a few examples of ways inhibition-related covariate effects (or the absence of such effects) can be confounded by processing speed effects. Growth rate effects can produce similar confounding. Therefore, failing to take different levels of context-independence violation into account can, in realistic scenarios, lead to qualitatively wrong conclusions despite the ordinal correspondence between true and non-parametric SSRT.

3.4. RDEX-ABCD as a measurement model

Given that non-parametric estimates of SSRT may cause misleading inferences, it is natural to ask whether the RDEX-ABCD model can be used as a measurement model to avoid such problems when analyzing ABCD data. To test whether RDEX-ABCD qualifies as a measurement model, we conducted parameter-recovery studies (methods described in Supplemental) using parameter estimates from our subsample to generate simulated individual data sets with same trial numbers and staircase algorithm as the empirical ABCD stop-signal data. We then used the same procedures as applied to the empirical data to estimate RDEX-ABCD parameters from the simulated data. As priors can impact parameter recovery, we repeated this procedure with both the informed and broad priors used with the empirical data. Figs. 8 and 9 display scatterplots of the relations between the data-generating and estimated parameters, associated correlation coefficients (r), and posterior "coverage" (c), the proportion of data-generating parameter values that fall within the 95 % posterior credible intervals provided by the Bayesian estimation procedure. If the coverage proportions are close to the nominal 95% value, this indicates that estimation is *calibrated* in terms of uncertainty.

As the points in Figs. 8 and 9 do not fall consistently higher or lower than the line indicating perfect recovery for any parameter, there is little evidence that parameters are systematically biased toward higher or lower values upon recovery. We also conducted more intensive parameter-recovery studies with a subset of individuals' parameter values that were intended to identify possible biases or second modes in parameter ranges (see Materials and Methods; results available at osf.io/2h8a7/), and again found little evidence for systematic bias. Coverage was nominal, or very close to nominal, in all cases, suggesting that inference based on Bayesian credible intervals will be well calibrated.

The parameters that characterize the go process and its probability of being triggered (v_+, v_-, B, t_0, and p_{gf}) are all recovered quite accurately ($r = 0.80$–0.97). Parameter estimates that characterize the stop process and its probability of being triggered (μ, σ, τ, p_{tf}) appear to be less precise or accurate, but generally show acceptable recovery, except for σ in the broad prior estimates, which is notably poor. Less accuracy and precision for the recovery of stop, relative to go, parameters is expected because stop trials are less frequent, and poorer recovery of these parameters occurs in standard paradigms for the BEESTS and RDEX models. Crucially, RDEX-ABCD recovers mean SSRT and trigger failure values well ($r = 0.75$–0.81). Given that 60 stop trials is barely above the minimal guideline for using non-parametric estimates in the standard stop-signal paradigm (Verbruggen et al., 2019) this represents excellent performance and suggests that the two key stop-related parameters estimated from RDEX-ABCD (SSRT, p_{tf}) can be used effectively in applied research.

Parameters for processing speed (v_0) and perceptual growth (g), which similarly depend on the sparse stop trial data, displayed poorer recovery. Therefore, if their values were of interest, a design in which participants performed many more trials would be required. However, as they are more likely to be considered nuisance parameters than of substantive interest in the ABCD data, and as there were no apparent biases in estimates of these parameters, this is unlikely to limit applications of the model. Furthermore, the coverage values indicate that, even if a subset of parameters' point estimates are not accurately recovered, the posterior distributions accurately reflect the uncertainty in these estimates and can therefore inform appropriately tentative inferences.

4. Discussion

We described a novel cognitive process modeling framework, RDEX-ABCD, that is aimed at illuminating the mechanistic processes that underlie performance on the ABCD stop-signal task. Critically, this framework accounts for unique context independence violations on the ABCD task due to a design feature in which the visual stop signal replaces the go choice stimulus, limiting the information participants need to make a choice. RDEX-ABCD integrates elements of prior parametric models of the stop-signal task (Logan et al., 2014; Matzke et al., 2013; Matzke et al., 2017; Tanis et al., 2022) with well-established accounts of masking effects on the processing of brief visual stimuli (Ratcliff and Rouder, 2000; Smith and Ratcliff, 2009; Smith and Sewell, 2013) by assuming that the ABCD design feature impacts the quality of evidence used for discrimination of go choice options. This account is sufficient to provide an excellent description of the pattern of increasing choice accuracy with greater SSD duration, which is the "smoking gun" that the task violates context independence (Bissett et al., 2021). Simulation studies that use RDEX-ABCD as the data generating model demonstrate that this violation can lead non-parametric SSRT estimates to reverse the ordering of inhibitory differences at both the individual and group levels. Fortunately, parameter-recovery studies indicate that RDEX-ABCD can be leveraged as a measurement model to avoid these problems and to reliably index SSRT, trigger failures, and other mechanistic processes of interest on the ABCD task.

Our exploration of comparison models within the RDEX-ABCD framework found that a simpler comparison model that assumed context independence could not account for lower choice accuracy at short SSDs, underscoring the need to address independence violations when modeling the ABCD task. However, we found that the simpler comparison model had little difficulty explaining other trends in ABCD stop trial data, including the inhibition function and increases in signal-respond RTs. This pattern of findings indicates that, although the context independence violation clearly affects the accuracy of go choices, its impact on the average latency of go choices at short SSDs may be relatively subtle. Such a dissociation may explain why RDEX-ABCD's SSRT estimates display a relatively strong ordinal correspondence with non-parametric SSRT. It also suggests that the independence violation has a more limited impact on non-parametric SSRT estimates than was originally feared (Bissett et al., 2021). Nonetheless, the simulation studies on SSRT biases show clear practical advantages to using parametric estimates from RDEX-ABCD relative to non-parametric methods, a point we return to below.

Although our model assumes that the ABCD task's context independence violation exclusively impacts the go choice process, Bissett

et al. (2021) speculated that the violation may also cause slowing in the stop process by "confusing" participants at short SSDs. We initially considered extensions to RDEX-ABCD that allow for impacts of the violation on the stop process but abandoned them due to difficulties with defining and estimating these impacts in ABCD data. Nonetheless, we think there are several reasons to believe that RDEX-ABCD's assumption that independence violations exclusively affect the go process provides a strong basis for ABCD applications. First, although the hypothesis that visual masking of the go choice stimulus degrades evidence quality is strongly indicated by a wealth of prior research (Kahneman, 1968; Ratcliff and Rouder, 2000; Smith and Ratcliff, 2009; Smith and Sewell, 2013), there is no such justification in prior research for a "confusion" account, making the possibility of impacts to the stop process speculative, and therefore a questionable assumption to include in a model. Second, the impact on go choice evidence quality assumed by RDEX-ABCD was sufficient to describe the pattern of choice accuracy across SSDs that is the context independence violation's key signature. Relevant to the principle of parsimony in modeling (Vandekerckhove et al., 2015), the absence of glaring misfits to this or other trends in stop trial data suggests that RDEX-ABCD provides an adequate account of the data generating process, and it is therefore unclear why additional parameters to account for impacts on the stop process would be needed. Finally, following the aphorism summarizing the message of Box (Box, 1976) that "all models are wrong, but some are useful", the RDEX-ABCD model displays clear practical utility by being both comprehensive enough to describe key trends in the data and parsimonious enough to allow for reliable measurement in the ABCD study. Future work, likely involving the collection of new data with greater numbers of stop trials, may be able to determine whether more complex models that allow for impacts of the violation on the stop process are better-supported than RDEX-ABCD. However, as models of greater complexity would likely be difficult to estimate in existing ABCD data, their practical utility would be questionable.

Given the unprecedented scientific opportunity afforded by the ABCD study and the importance of unbiased SSRT estimates for researching inhibitory ability, our findings have several key implications. First, they suggest that major changes to the ABCD task design are not warranted, ensuring longitudinal comparability of the behavioral and fMRI data between waves. SSRT measurement issues related to the violation appear to be manageable using RDEX-ABCD, and the practical difficulties of using the model are relatively minor compared to the problems introduced by breaking longitudinal comparability. Furthermore, even if alternate models that explain the context-independence violation with different processes are proposed and supported, our results provide a general demonstration that cognitive modeling can effectively overcome limitations related to the ABCD design.

The implications of our work for analyses of already-collected ABCD data are nuanced. As we found that non-parametric SSRT estimates calculated using recommended best practices (Verbruggen et al., 2019) generally preserved the rank ordering, if not the absolute values, of participants' SSRT, it is possible that inferences based on the non-parametric estimates may not be misleading in many situations. However, we also showed that non-parametric estimates can lead to incorrect (including reversed) inferences when a parameter that explains the context independence violations is confounded with a covariate of interest. It seems plausible that such confounding might occur in practice. Rather than taking the chance of assuming that such confounding is not present, we recommend that researchers use parametric measurement models, such as the one we propose, that account for the context-independence violations evident in the ABCD data.

We note that this trade-off between the precision of cognitive process modeling and the ease of using non-parametric SSRT estimates is not unique to ABCD. As outlined in the introduction, trigger failures in standard designs cannot be easily accommodated using non-parametric methods and have already been shown to bias SSRT estimates (Matzke et al., 2017) and distort substantive conclusions (Matzke et al., 2017; Weigard et al., 2019) if ignored. Indeed, current consensus recommendations for estimating SSRT from the stop-signal task (Verbruggen et al., 2019) acknowledge that cognitive process models, despite being difficult for researchers with less technical expertise to implement, provide less biased estimates of SSRT relative to even the best non-parametric methods. Our findings suggest parallel recommendations for analysis of the ABCD task; whenever it is feasible to do so, validated cognitive process models such as RDEX-ABCD should be used to estimate SSRT, trigger failure, and other mechanistic processes while accounting for ABCD-specific context-independence violations.

We are taking two steps to facilitate the wider adoption of RDEX-ABCD. First, we have shared, on the Open Science Framework (osf.io/2h8a7), the code we used to specify and fit the model within Dynamic Models of Choice (DMC), a free set of R functions for Bayesian estimation of evidence accumulation models that comes equipped with comprehensive, hands-on tutorials for first-time users (Heathcote et al., 2019). Researchers can now freely use our code to estimate parameters of the RDEX-ABCD model using identical procedures to those implemented in the current study or flexibly alter the model and fitting procedures to suit their specific research aims. Second, we are now working towards the goal of sharing RDEX-ABCD model parameter estimates for the entire ABCD sample, either as a new NDA collection or in future ABCD Study data releases, as soon as possible.

One notable feature of our analysis strategy is that we used priors generated from a hierarchical model fit to an independent subsample of ABCD participants to inform individual-level estimation. This method provides the key benefit of a fully hierarchical approach (prevention of over-fitting to individual-level data by using information about group distributions) without two of this approach's drawbacks. The first drawback is that fitting hierarchical models to very large data sets is demanding in terms of computational resources and technical expertise, although it has been done for other large-scale projects using simpler models (PISA, 2018). Therefore, to simplify the computational demands needed to fit the model, we focused on individual-analysis methods that require only a modern multi-core PC, so that the benefits of RDEX-ABCD are more immediately available. That said, we believe that developing methods for fitting large-scale hierarchical cognitive process models is a worthwhile aim for not only the stop-signal task, but also the other tasks used in the ABCD project. Given the large number of sampling chains required by the DE-MCMC method (Turner et al., 2013), a promising strategy for reducing the computational demands of such models is to use recently developed samplers that are more efficient, such as particle Metropolis within Gibbs sampling (Gunawan et al., 2020).

Even if these computational limitations are addressed, however, a second drawback of fully hierarchical models is that individual-level estimates drawn from them are unsuitable for follow-up frequentist or Bayesian inferential methods because they may reduce between-subjects variability in parameter estimates in a way that violates the independence assumptions of the inferential tests, biasing them toward finding effects (Boehm et al., 2018; Evans and Wagenmakers, 2019). Informed priors also constrain variability in parameter estimates but, as estimation is carried out independently for each participant, this approach is better aligned with the assumptions of follow-up frequentist tests. "Plausible values" analyses have been proposed as a solution for bivariate correlations with hierarchical cognitive model parameters (Ly et al., 2017), but until this approach is extended to the multi-level modeling, structural equation modeling, and multivariate prediction methods commonly used with ABCD data (which is another critical area for future work), individual-level estimation with informed priors provides a good alternative solution. That said, simulation studies to assess Type 1 error rates in each application or sensitivity analyses involving parameter estimates from both broad and informative priors can be used to rule out the possibility the priors bias inferences.

A key limitation RDEX-ABCD shares with the original RDEX model is the fact that, although it allows specific mechanisms that impact the go process to be measured (evidence accumulation rates, thresholds), the

outcome of the stop process is modeled with ex-Gaussian distributional parameters, which cannot be used to measure such specific mechanisms (Matzke and Wagenmakers, 2009). As is demonstrated by the array of stop-signal task models outlined in the introduction, cognitive process models exist on a continuum that ranges from simpler "descriptive" models that posit coarse, summary-level descriptions of the processes involved (e.g., the non-parametric race model) to those that posit specific parameters that provide a more complete explanation of how these processes operate (e.g., the racing-diffusion model). Although the racing-diffusion model (Logan et al., 2014) dissociates mechanistic parameters for both the go and stop processes, the parameter recovery problems this model exhibits prevents it from being used to measure these mechanisms. Therefore, we argue that the RDEX framework currently provides a valuable "middle ground" along this continuum by allowing for more specific features of the stop process to be dissociated than are indexed by the original race model (latency, variability, and trigger failure) while simultaneously offering excellent measurement properties.

In summary, we introduce a cognitive process modeling framework that explains the impact of context-independence violations on the ABCD Study's stop-signal task and, in doing so, accounts for key trends in the ABCD data. We show that failing to account for context-independence violations could produce misleading inferences, and that the proposed model provides a practical remedy, enabling unbiased and reliable estimation of SSRT and other key process parameters that contribute to task performance. We argue that the model can advance ABCD Study research efforts by improving the measurement of inhibition and other cognitive processes (e.g., trigger failure and choice evidence accumulation) with existing ABCD stop-signal task data. More broadly, this work highlights the critical strengths of a cognitive process modeling approach for increasing the precision of both theories and measures of neurocognitive phenomena.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

A.W. was supported by National Institute on Alcohol Abuse & Alcoholism T32 AA007477 (to Frederic Blow) and National Institute on Drug Abuse K23 DA051561.

DM was supported by a Vidi grant (VI.Vidi.191.091) from the Netherlands Organization for Scientific Research (NWO). C.T. was supported by an Advanced ERC grant 743086 UNIFY (to Eric-Jan Wagenmakers). A.H. was supported by 743086 UNIFY, Australian Research Council grants DP210100313 and DP200100655, and a Révész Visiting Professor Fellowship at the Department of Psychology, University of Amsterdam.

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development[SM] (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multi-site, longitudinal study designed to recruit more than 10,000 children age 9–10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from ABCD Release 1.1, DOI 10.15154/1412097.

Data statement

Data analyzed in this manuscript were from the Adolescent Brain Cognitive Development (ABCD) Study and are available through the NIH Data Archive (NDA): nda.nih.gov/abcd. Code files for all analyses featured in this study are openly available on the Open Science Framework (OSF): osf.io/2h8a7/.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dcn.2022.101191.

References

Ando, T., 2011. Predictive Bayesian model selection. Am. J. Math. Manag. Sci. 31 (1–2), 13–38.
Aron, A.R., Poldrack, R.A., 2006. Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. J. Neurosci. 26 (9), 2424–2433.
Band, G.P.H., van der Molen, M.W., Logan, G.D., 2003. Horse-race model simulations of the stop-signal procedure. Acta Psychol. 112 (2), 105–142. https://doi.org/10.1016/S0001-6918(02)00079-3.
Bissett, P.G., Hagen, M.P., Jones, H., Poldrack, R.A., 2021. Design issues and solutions for stop-signal data from the adolescent brain cognitive development (ABCD) study. eLife.
Boehm, U., Marsman, M., Matzke, D., Wagenmakers, E.-J., 2018. On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. Behav. Res. Methods 50 (4), 1614–1631.
Boonstra, A.M., Kooij, J., Oosterlaan, J., Sergeant, J.A., Buitelaar, J.K., 2010. To act or not to act, that's the problem: primarily inhibition difficulties in adult ADHD. Neuropsychology 24 (2), 209.
Box, G., 1976. Science and statistics. J. Am. Stat. Assoc. 71, 791–799.
Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14 (5), 365–376.
Casey, B., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., others, 2018. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54.
Clark, D.A., Hicks, B.M., Angstadt, M., Rutherford, S., Taxali, A., Hyde, L.W., Weigard, A., Heitzeg, M.M., Sripada, C., 2021. The general factor of psychopathology in the adolescent brain cognitive development (ABCD) study: a comparison of alternative modeling approaches. Clin. Psychol. Sci.
Colonius, H., Özyurt, J., Arndt, P.A., 2001. Countermanding saccades with auditory stop signals: Testing the race model. Vis. Res. 41 (15), 1951–1968.
Dick, A.S., Garcia, N.L., Pruden, S.M., Thompson, W.K., Hawes, S.W., Sutherland, M.T., Riedel, M.C., Laird, A.R., Gonzalez, R., 2019. No evidence for a bilingual executive function advantage in the ABCD study. Nat. Hum. Behav. 3 (7), 692–701.
Donkin, C., Brown, S.D., 2018. Response times and decision-making. Stevens' Handb. Exp. Psychol. Cogn. Neurosci. 5, 1–33.
Efron, B., Morris, C., 1977. Stein's paradox in statistics. Sci. Am. 236 (5), 119–127.
Etz, A., Vandekerckhove, J., 2016. A Bayesian perspective on the reproducibility project: psychology. PLOS One 11 (2), e0149794.
Evans, N.J., Wagenmakers, E.-J., 2019. Theoretically meaningful models can answer clinically relevant questions. Brain 142 (5), 5.
Falk, E.B., Hyde, L.W., Mitchell, C., Faul, J., Gonzalez, R., Heitzeg, M.M., Keating, D.P., Langa, K.M., Martz, M.E., Maslowsky, J., others, 2013. What is a representative brain? Neuroscience meets population science. Proc. Natl. Acad. Sci. USA 110 (44), 17615–17622.
Funkhouser, C.J., Chacko, A.A., Correa, K.A., Kaiser, A.J., Shankman, S.A., 2020. Unique longitudinal relationships between symptoms of psychopathology in youth: a cross-lagged panel network analysis in the ABCD study. J. Child Psychol. Psychiatry.
Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., Zahs, D., 2018. Recruiting the ABCD sample: design considerations and procedures. Dev. Cogn. Neurosci. 32, 16–22.
Gelman, A., Rubin, D.B., others, 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7 (4), 457–472.
Gelman, A., Meng, X.-L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Stat. Sin. 733–760.
Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. , 2013. Bayesian data analysis. CRC press.

Gorenstein, E.E., Newman, J.P., 1980. Disinhibitory psychopathology: a new perspective and a model for research. Psychol. Rev. 87 (3), 301.
Gunawan, D., Hawkins, G.E., Tran, M.-N., Kohn, R., Brown, S., 2020. New estimation approaches for the hierarchical linear ballistic accumulator model. J. Math. Psychol. 96, 102368.
Heathcote, A., Popiel, S.J., Mewhort, D., 1991. Analysis of response time distributions: an example using the Stroop task. Psychol. Bull. 109 (2), 340.
Heathcote, A., Brown, S.D., Wagenmakers, E.-J., 2015. An introduction to good practices in cognitive modeling. An Introduction to Model-based Cognitive Neuroscience. Springer, pp. 25–48.
Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., Matzke, D., 2019. Dynamic models of choice. Behav. Res. Methods 51 (2), 961–985.
Kahneman, D., 1968. Method, findings, and theory in studies of visual masking. Psychol. Bull. 70 (6P1).
Lee, M.D., Wagenmakers, E.-J., 2014. Bayesian Cognitive Modeling: A Practical Course. Cambridge university press.
Lipszyc, J., Schachar, R., 2010. Inhibitory control and psychopathology: a meta-analysis of studies using the stop signal task. J. Int. Neuropsychol. Soc.: JINS 16 (6), 1064.
Logan, G.D. , 1994. On the ability to inhibit thought and action: A users' guide to the stop signal paradigm.
Logan, G.D., Cowan, W.B., 1984. On the ability to inhibit thought and action: a theory of an act of control. Psychol. Rev. 91 (3), 295.
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. J. Exp. Psychol.: Hum. Percept. Perform. 10 (2), 276–291. https://doi.org/10.1037/0096-1523.10.2.276.
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. Psychol. Rev. 121 (1), 66.
Loken, E., Gelman, A., 2017. Measurement error and the replication crisis. Science 355 (6325), 584–585.
Ly, A., Boehm, U., Heathcote, A., Turner, B.M., Forstmann, B., Marsman, M., Matzke, D., 2017. A flexible and efficient hierarchical Bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. Comput. Models Brain Behav. 467–480.
Mahmood, O., Goldenberg, D., Thayer, R., Migliorini, R., Simmons, A., Tapert, S., 2013. Adolescents' fMRI activation to a response inhibition task predicts future substance use. Addict. Behav. 38 (1), 1435–1441.
Marek, S., Tervo-Clemmens, B., Nielsen, A.N., Wheelock, M.D., Miller, R.L., Laumann, T. O., Earl, E., Foran, W.W., Cordova, M., Doyle, O., others, 2019. Identifying reproducible individual differences in childhood functional brain networks: an ABCD study. Dev. Cogn. Neurosci. 40, 100706.
Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Donohue, M.R., Foran, W., Miller, R.L., Hendrickson, T.J., others, 2022. Reproducible brain-wide association studies require thousands of individuals. Nature 1–7.
Matzke, D., Wagenmakers, E.-J., 2009. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. Psychon. Bull. Rev. 16 (5), 5. https://doi.org/10.3758/PBR.16.5.798.
Matzke, D., Dolan, C.V., Logan, G.D., Brown, S.D., Wagenmakers, E.-J., 2013. Bayesian parametric estimation of stop-signal reaction time distributions. J. Exp. Psychol.: Gen. 142 (4), 1047.
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. Atten. Percept. Psychophys. 79 (4), 1078–1086.
Matzke, D., Love, J., Heathcote, A., 2017. A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. Behav. Res. Methods 49 (1), 267–281.
Matzke, D., Verbruggen, F., Logan, G.D., 2018. The stop-signal paradigm. Stevens' Handb. Exp. Psychol. Cogn. Neurosci. 5, 1–45.
Matzke, D., Curley, S., Gong, C.Q., Heathcote, A., 2019. Inhibiting responses to difficult choices. J. Exp. Psychol.: Gen. 148 (1), 124.
Matzke, D., Logan, G.D., Heathcote, A., 2020. A cautionary note on evidence-accumulation models of response inhibition in the stop-signal paradigm. Comput. Brain Behav. https://doi.org/10.1007/s42113-020-00075-x.
Matzke, D., Strickland, L.J.G., Sripada, C., Weigard, A.S., Puri, R., He, J., Hirst, R., & Heathcote, A. , 2021. Stopping timed actions.
Mazurek, M.E., Roitman, J.D., Ditterich, J., Shadlen, M.N., 2003. A role for neural integrators in perceptual decision making. Cereb. Cortex 13 (11), 1257–1269.
Mennies, R.J., Birk, S.L., Norris, L.A., Olino, T.M., 2020. The main and interactive associations between demographic factors and psychopathology and treatment utilization in youth: a test of intersectionality in the ABCD study. J. Abnorm. Child Psychol. 1–13.
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. Cogn. Psychol. 41 (1), 49–100.
Nigg, J.T., 2017. Annual research review: on the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. J. Child Psychol. Psychiatry 58 (4), 361–383.
Nigg, J.T., Wong, M.M., Martel, M.M., Jester, J.M., Puttler, L.I., Glass, J.M., Adams, K.M., Fitzgerald, H.E., Zucker, R.A., 2006. Poor response inhibition as a predictor of problem drinking and illicit drug use in adolescents at risk for alcoholism and other substance use disorders. J. Am. Acad. Child Adolesc. Psychiatry 45 (4), 468–475.
PISA , 2018. Technical Report, Chapter 9. https://www.oecd.org/pisa/sitedocument/PISA-2015-Technical-Report-Chapter-9-Scaling-PISA-Data.pdf.
Pitt, M.A., Myung, I.J., 2002. When a good fit can be bad. Trends Cogn. Sci. 6 (10), 421–425.
Poldrack, R.A., Gorgolewski, K.J., 2014. Making big data open: data sharing in neuroimaging. Nat. Neurosci. 17 (11), 1510–1517.
Ratcliff, R., Rouder, J.N., 2000. A diffusion model account of masking in two-choice letter identification. J. Exp. Psychol.: Hum. Percept. Perform. 26 (1), 127.
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 Cogn. 56 (2), 129–140.
Skippen, P., Matzke, D., Heathcote, A., Fulham, W.R., Michie, P., Karayanidis, F., 2019. Reliability of triggering inhibitory process is a better predictor of impulsivity than SSRT. Acta Psychol. 192, 104–117.
Smith, P.L., Little, D.R., 2018. Small is beautiful: in defense of the small-N design. Psychon. Bull. Rev. 25 (6), 2083–2101.
Smith, P.L., Ratcliff, R., 2009. An integrated theory of attention and decision making in visual signal detection. Psychol. Rev. 116 (2), 283.
Smith, P.L., Sewell, D.K., 2013. A competitive interaction theory of attentional selection and decision making in brief, multielement displays. Psychol. Rev. 120 (3), 589.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc.: Ser. b (Stat. Methodol.) 64 (4), 583–639.
Sripada, C., Rutherford, S., Angstadt, M., Thompson, W.K., Luciana, M., Weigard, A., Hyde, L.H., Heitzeg, M., 2019. Prediction of neurocognition in youth from resting state fMRI. Mol. Psychiatry 1–9. https://doi.org/10.1038/s41380-019-0481-6.
Tanis, C., Heathcote, A., Zrubka, M., & Matzke, D. , 2022. A hybrid approach to dynamic cognitive psychometrics. PsyArXiv.
Tannock, R., Schachar, R., Logan, G., 1995. Methylphenidate and cognitive flexibility: dissociated dose effects in hyperactive children. J. Abnorm. Child Psychol. 23 (2), 235–266.
Tillman, G., Van Zandt, T., Logan, G.D., 2020. Sequential sampling models without random between-trial variability: the racing diffusion model of speeded decision making. Psychon. Bull. Rev. 27, 911–936.
Turner, B.M., Sederberg, P.B., Brown, S.D., Steyvers, M., 2013. A method for efficiently sampling from distributions with correlated dimensions. Psychol. Methods 18 (3), 368.
van Ravenzwaaij, D., Brown, S.D., Marley, A., Heathcote, A., 2019. Accumulating advantages: a new conceptualization of rapid multiple choice. Psychol. Rev.
Vandekerckhove, J., Matzke, D., Wagenmakers, E.-J., 2015. Model comparison and the principle. Oxf. Handb. Comput. Math. Psychol. 300.
Verbruggen, F., Logan, G.D., 2008. Response inhibition in the stop-signal paradigm. Trends Cogn. Sci. 12 (11), 418–424.
Verbruggen, F., Chambers, C.D., Logan, G.D., 2013. Fictitious inhibitory differences: how skewness and slowing distort the estimation of stopping latencies. Psychol. Sci. 24 (3), 352–362.
Verbruggen, F., McLaren, I.P., Chambers, C.D., 2014. Banishing the control homunculi in studies of action control and behavior change. Perspect. Psychol. Sci. 9 (5), 497–524.
Verbruggen, F., Aron, A.R., Band, G.P., Beste, C., Bissett, P.G., Brockett, A.T., Brown, J. W., Chamberlain, S.R., Chambers, C.D., Colonius, H., others, 2019. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. eLife 8, e46323.
Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., Pérez-Stable, E.J., Riley, W.T., Bloch, M.H., Conway, K., others, 2018. The conception of the ABCD study: from substance use to a broad NIH collaboration. Dev. Cogn. Neurosci. 32, 4–7.
Weigard, A., Heathcote, A., Matzke, D., Huang-Pollock, C., 2019. Cognitive modeling suggests that attentional failures drive longer stop-signal reaction time estimates in attention deficit/hyperactivity disorder. Clin. Psychol. Sci. 7 (4), 856–872. https://doi.org/10.1177/2167702619838466.