Comparing One-Boundary and Two-Boundary Evidence Accumulation Models for Go/No-Go Processes: An Application to the Decision to Shoot

Yiyang Chen (ych17@iu.edu)
Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Jun Fang (junfang@iu.edu)
Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Joseph Cesario (cesario@msu.edu)
Department of Psychology, 316 Physics Rd
East Lansing, MI 48824

Taosheng Liu (tsliu@msu.edu)
Department of Psychology, 316 Physics Rd
East Lansing, MI 48824

Tim Pleskac (tpleskac@iu.edu)
Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Abstract

The cognitive processes underlying Go/No-Go performance may be explained by two plausible evidence accumulation models: Two-Boundary (2-B) and One-Boundary (1-B) decision drift models (DDMs). While both embed a Go decision, the 2-B DDM embeds a definitive No-Go decision, whereas the 1-B DDM embeds a response window for Go. Using simulations, we found that model comparison methods like leave-one-out cross-validation (LOO), coupled with Bayesian hierarchical modeling, can correctly identify the underlying model. Additionally, using the correct model reduces the risk of missing true effects or detecting spurious findings. Therefore, we recommend researchers implement and compare both models for Go/No-Go studies to reduce misleading results. Lastly, we implemented these models to investigate race effects in the decision to shoot during police training. We found that the accumulated evidence needed to reach the Shoot decision is lower for Black suspects, which explains the heightened error rates for shooting unarmed Black suspects in data.

Keywords: Go/No-Go; Response Time; Evidence Accumulation; Police Decision-Making; One-Boundary; Two-Boundary.

Introduction

The Go/No-Go task (Donders, 1969) is one of the most commonly used paradigms in psychological assessment. Serving as a strong tool for measuring cognition and decision-making, the Go/No-Go paradigm has been applied to various fields including but not limited to attention (Meule, 2017), response inhibition (Wright, Lipszyc, Dupuis, Thayapararajah, & Schachar, 2014), social cognition (Nosek & Banaji, 2001), decision making (Ba, Zhang, Salvendy, Cheng, & Ventsislavova, 2016), brain activities and neural correlates (Eagle, Bari, & Robbins, 2008; Rubia et al., 2001), and clinical assessment (Rubia, Smith, & Taylor, 2007). In a typical Go/No-Go experiment, participants are required to report a response to one of two choices but withhold the response to the other choice. As a result, observed responses and their corresponding response times (RTs) are available for only a proportion of trials in the Go/No-Go paradigm.

To date, researchers have introduced several evidence accumulation models to explain the processes underlying Go/No-Go responses and RTs (Gomez, Ratcliff, & Perea, 2007; Ratcliff, Huang-Pollock, & McKoon, 2018; Myers et al., 2023; Trueblood, Endres, Busemeyer, & Finn, 2011). Broadly, these models can be divided into two categories: Two-Boundary (2-B) and One-Boundary (1-B). These models capture two distinct response mechanisms in the task. The 2-B models embed the assumption that the Go/No-Go task involves choosing between two distinct options: a Go decision that produces a response, and a No-Go decision that withholds a response. For example, the decision diffusion model (DDM) from Gomez et al. (2007) posits that participants accumulate evidence supporting either a Go or a No-Go response. When the evidence reaches the Go boundary, a Go response is executed; conversely, if the evidence instead reaches the No-Go boundary, the response is withheld. In contrast, 1-B models assume that only a Go Decision is actively considered, while the absence of a decision (i.e., not reaching the boundary in time) functions as the No-Go response. Under this framework, a single process accumulates evidence for a Go response within a certain response window. If the Go boundary is reached before a time-out threshold is reached (which determines the duration of the response window), the participant responds; otherwise, no response is made. This time-out threshold may be determined exogenously (e.g., by the experimenter or some other external factors) or endogenously (i.e., by the participant).

Although the 1-B and 2-B approaches represent theoretically distinct response mechanisms, to date, there has only been one published direct comparison of these different response mechanisms (Gomez et al., 2007). They found that across a range of common cognitive psychology tasks (visual word recognition and numerosity discrimination), the 2-B DDM best explained choice and response time data for Go/No-Go tasks. We contend, however, that the 2-B model may not be a general response process across all Go/No-Go Tasks. For instance, when the experimental stimuli are simple, static, and unambiguous, participants may be more likely to follow a two-boundary mechanism with a definitive No-Go decision. In these cases, apart from the unobserved No-Go responses, these tasks are not fundamentally different from their two-choice variants (Ratcliff, 1978; Brown & Heathcote, 2008). For instance, in a recognition decision, it seems quite plausible that participants map a Go/No-Go response to an Old vs New response. However, other tasks may not align as closely to their two-choice variants. For example, during a police officer's decision to shoot, it may be difficult for participants to make a No-Go decision (i.e., deem the suspect as harmless) when they face dynamic, naturalistic scenarios (Pleskac, Cesario, Johnson, & Gagnon, 2024), as the suspects could drastically change their behaviors in the next scene.

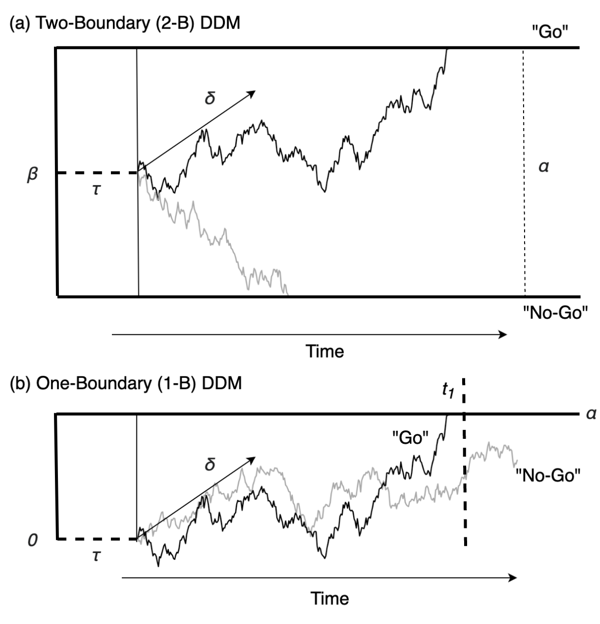

Figure 1: Model structures for 2-B and 1-B DDMs.

We sought to develop a Bayesian computational modeling solution for identifying the different response processes in Go/No-Go tasks. Hence, we developed a Bayesian hierarchical implementation of these two Go/No-Go accumulation models, namely 2-B DDM and 1-B DDM. We implemented them in STAN. We then used simulations to examine three aspects of these two models: (i) Whether we can identify the true mechanism (accurate model) underlying data using Bayesian model evaluation methods, such as leave-one-out cross-validation (LOO) (Vehtari, Gelman, & Gabry, 2017); (ii) whether we can identify true effects in data, such as race effects in the decision to shoot (Mekawi & Bresin, 2015), when the accurate model (consistent with the underlying mechanism) is used to analyze data; (iii) whether we can still identify true effects in data when the wrong model (inconsistent with the underlying mechanism) is used to analyze data. We then implemented the models to an empirical Go/No-Go dataset assessing the decision-making processes of cadets in a police academy. We investigated whether the suspect's race (Black or White) impacted the participants' decisions to use deadly force. Lastly, based on our findings, we provided recommendations for Go/No-Go modeling and discussed the limitations of this study.

RT models and Bayesian hierarchical implementation

The 2-B and 1-B DDMs' structures are shown in Figure 1. Because we later implemented the models to understand race effects during the decision to shoot (Pleskac et al., 2024), we framed the models using the police decision-making framework. However, these models can be easily re-framed for other studies. During the task, participants are asked to make a Shoot (Go) response when the suspect presents a gun, and withhold a Shoot response (No-Go) when the suspect presents a harmless object. We considered scenarios in which the suspects are White or Black.

2-B DDM with definitive No-Go decision

The 2-B DDM is shown in Figure 1(a). We implemented this model using custom code in the R package "rstan" (Stan Development Team et al., 2020). The upper decision boundary corresponds to the Go decision, while the lower decision boundary corresponds to the No-Go decision. When participants have accumulated enough evidence to reach the upper/lower boundaries, the Go/No-Go decisions are made, respectively. For Go observations, the RTs follow the Wiener distribution, and we modeled them using the built-in 4-parameter Wiener distribution in STAN. For No-Go cases, we modeled their probability of occurrences as the probability of hitting the lower boundary, whose closed form is provided by Ratcliff and Tuerlinckx (2002) and Wang and Busemeyer (2021).

The separation of decision boundaries is characterized by the boundary separation parameter α ($\alpha > 0$). The parameter τ denotes non-decision time. The starting point parameter β ($0 < \beta < 1$) reflects tendencies in response. Participants become more likely to reach a Go decision as β approaches 1. Lastly, noisy evidence is being accumulated, and the drift rate δ characterizes the average speed of evidence accumulation. Participants are more likely to accumulate evidence endorsing the Go decision when $\delta > 0$. We implemented the 2-B DDM using Bayesian hierarchical methods. Models in this study were designed for a group of N participants each performing I trials, thus we used n to denote participant number and i as trial number.

In the decision-to-shoot example, for Participant n, the decision boundary α_n and starting point β_n are pre-determined in the decision process before knowing whether the suspect has a gun, thus they are affected by the suspect's race only:

$$\alpha_n = \exp(\alpha_{1,n}\text{White}_{i,n} + \alpha_{2,n}\text{Black}_{i,n}),$$
$$\beta_n = \text{inv.logit}(\beta_{1,n}\text{White}_{i,n} + \beta_{2,n}\text{Black}_{i,n}), \quad (1)$$

where $\text{White}_{i,n}$ and $\text{Black}_{i,n}$ are binary indicators identifying whether the suspect shown to Participant n on Trial i is White or Black, respectively. The drift rate δ_n is affected by both gun presence and the suspect's race,

$$\delta_n = 8 \cdot \text{inv.logit}(\delta_{1,n}\text{NG}_{i,n}\text{White}_{i,n} + \delta_{2,n}\text{NG}_{i,n}\text{Black}_{i,n}$$
$$\delta_{3,n}\text{G}_{i,n}\text{White}_{i,n} + \delta_{4,n}\text{G}_{i,n}\text{Black}_{i,n}) - 4, \quad (2)$$

where $\text{G}_{i,n}$ and $\text{NG}_{i,n}$ are binary identifiers for gun and non-gun conditions. We also made the restriction $-4 < \delta_n < 4$ to avoid potential technical issues during model implementation. We assumed that Participant n had a non-decision time of τ_n and a decision boundary of α_n that are fixed across trials. The Bayesian implementation is described later in "Bayesian hierarchical implementation in STAN."

1-B DDM with response window

The 1-B DDM is shown in Figure 1(b). We implemented this model using custom code in "rstan." The 1-B model features only one decision boundary corresponding to the Go decision, along with a time-out threshold that determines the duration of the response window. If accumulated evidence reaches the boundary prior to the threshold, a Go response is observed; otherwise No-Go is observed. Because the RT distribution for 1-B diffusion processes follows the Wald distribution (Matzke, Logan, & Heathcote, 2020; Trueblood et al., 2011; Steingroever, Wabersich, & Wagenmakers, 2021), we specified a custom-written Wald distribution function for the RTs under Go responses. We specified the probability of No-Go observations to follow the complementary Wald CDF, which reflects the probability of accumulated evidence hitting the boundary later than the time-out threshold.

Using the decision-to-shoot framework, the 1-B DDM is specified as follows. For Participant n, we denoted the boundary parameter as α_n, which is impacted by suspect race like Equation (1), and we denoted the drift rate as δ_n, which is impacted by both gun presence and race like Equation (2) but with the restriction $0 < \delta_n < 4$. We assumed that Participant n had a non-decision time of τ_n and a response threshold of $t_{1,n}$ that are fixed across trials.

Bayesian hierarchical implementation in STAN

We implement the models in Bayesian hierarchical frameworks using STAN and rstan. In both the 2-B and 1-B DDMs, for individual-level parameters that embed effects from race or gun presence (including $\alpha_{1,n}$, $\alpha_{2,n}$, $\beta_{1,n}$, $\beta_{2,n}$, $\delta_{1,n}$, $\delta_{2,n}$, $\delta_{3,n}$, and $\delta_{4,n}$), if we use θ_n to symbolize all these parameters, the individual-level priors for them are $\theta_n \sim N(\theta, .5)$. In this study, we mainly examined the conditions in which each participant performed limited trials (<50 trials), thus we fixed the standard deviation in these priors as .5 and only allowed group-level means to be estimated[1]. The hyperpriors for group-level means are $\theta \sim N(0,1)$, which is consistent across all aforementioned parameters.

In both models, we set Participant n's upper limit for non-decision time as $T_{0,n}$ where $T_{0,n}$ is determined by the lowest observed response time ($T_{0,n} = \min(.99 \times \min(RT), 0.4)$). The prior for the non-decision time τ_n is set as $\text{logit}(\frac{\tau_n}{T_{0,n}}) \sim N(0,1)$. For the 1-B DDM, we treated the time-out thresholds as endogenously determined and hence estimated it for each participant. We set Participant n's lower limit for the time-out threshold as $T_{1,n}$, which is calculated as the largest observed RT plus 0.01 seconds. The prior for the time-out threshold $t_{1,n}$ is set as $\log(t_{1,n} - T_{1,n}) \sim N(0,1)$.

Simulation Study

In this simulation, we aimed to examine whether and how often the accurate model (1-B or 2-B) can be identified from data using LOO[2], whether true race effects can be identified using the accurate model, and whether these race effects can be identified using the wrong model. We first described the simulated data and then reviewed the results related to each question of interest.

Data simulation and model implementation

We simulated the scenarios in which participants make the decision to shoot/not-shoot depending on whether the suspect presents a gun or a harmless object. We simulated race effects depending on whether the suspect is White or Black. We designed four conditions under the 2-B DDM, and another four conditions under the 1-B DDM. We generated 50 simulated datasets for each condition, where each dataset included 50 participants performing 40 trials each. We selected these numbers because they were relatively consistent with our empirical dataset.

In the 2-B DDM conditions (Table 1), we simulated: (1) Condition "2-B: No effect", which embedded no race effects; (2) Condition "2-B: drift NG", which embedded a race difference in the drift rates for non-gun conditions; (3) Condition "2-B: drift G", which embedded a race difference in the drift rates for gun conditions; and (4) Condition "2-B: start point", which embedded a race difference in the starting points. Parameters not specified in Table 1 had the following distributions: $\delta_{1,n}$ and $\delta_{2,n}$ followed $N(-1, .5)$, $\delta_{3,n}$ and $\delta_{4,n}$ followed $N(1, .5)$, $\beta_{1,n}$ and $\beta_{2,n}$ followed $N(0, .5)$, $\alpha_{1,n}$ and $\alpha_{2,n}$ followed $N(1, .1)$, and τ_n followed $U(.1, .4)$. These distributions

[1]However, we recommend setting the standard deviations as parameters when the empirical data includes sufficient trials per participant (Pleskac, Cesario, & Johnson, 2018; Pleskac et al., 2024).

[2]Model comparison results using the widely applicable information criterion (WAIC) (Vehtari et al., 2017) are consistent with results using LOO.

Table 1: Data generating parameters for race effects.

	White			Black		
Condition	Param	M	SD	Param	M	SD
2-B: No effect	-	-	-	-	-	-
2-B: drift NG	$\delta_{1,n}$	-.5	.5	$\delta_{2,n}$	-1	.5
2-B: drift G	$\delta_{3,n}$	.5	.5	$\delta_{4,n}$	1	.5
2-B: start point	$\beta_{1,n}$	-.25	.5	$\beta_{2,n}$	.25	.5
1-B: No effect	-	-	-	-	-	-
1-B: drift NG	$\delta_{1,n}$	-4	1	$\delta_{2,n}$	-2	1
1-B: drift G	$\delta_{3,n}$	0	.5	$\delta_{4,n}$	.5	.5
1-B: boundary	$\alpha_{1,n}$	.5	.5	$\alpha_{2,n}$	1	.5

were selected to reflect plausible values in empirical studies.

In the 1-B DDM conditions (Table 1), we simulated: (1) Condition “1-B: No effect”, which embedded no race effects; (2) Condition “1-B: drift NG”, which embedded a race difference in the drift rates for non-gun conditions; (3) Condition “1-B: drift G”, which embedded a race difference in the drift rates for gun conditions; and (4) Condition “1-B: boundary”, which embedded a race difference in the decision boundaries. Parameters not specified in Table 1 had the following distributions: $\delta_{1,n}$ and $\delta_{2,n}$ followed $N(-4,1)$, $\delta_{3,n}$ and $\delta_{4,n}$ followed $N(0,.5)$, $\alpha_{1,n}$ and $\alpha_{1,n}$ followed $N(.5,.5)$, τ_n followed $U(.1,.4)$, and the RT thresholds $t_{1,n}$ were randomly sampled between 2 and 3.

After generating the simulated datasets for all eight conditions, we then applied both Bayesian hierarchical 1-B and 2-B DDMs to each dataset using “rstan,” obtaining one chain with 500 warm-ups and 4000 samples each.

Model selection results

In all conditions, we found that LOO was able to select the accurate, data-generating model in 100% cases across the 50 simulations that we investigated. Therefore, using Bayesian hierarchical modeling, for these experimental conditions, it is likely that we can identify the mechanism (1-B or 2-B DDM).

Race effects using accurate models

We then examined whether the accurate models consistent with data-generating mechanisms can successfully recover the race effects embedded in each condition. Table 2 displays the results. For each set of estimated parameters $\hat{\theta}_1^{(k)}$ and $\hat{\theta}_2^{(k)}$ ($k = 1, 2, ..., 4000$ samples), we displayed the numbers of simulations in which $p(\hat{\theta}_1^{(k)} < \hat{\theta}_2^{(k)}) > .95$ (shown in columns “$<$”) and $p\hat{\theta}_1^{(k)} > \hat{\theta}_2^{(k)}) > .95$ (shown in columns “$>$”). Cases embedding true race effects are boldfaced, while those without true race effects are not boldfaced.

Based on Table 2, race effects are likely to be recovered by the accurate model. Using the 2-B model to fit 2-B data, we identified true race effects in 84%-100% cases, and found spurious race effects in at most 16% cases. Using the 1-B model to fit 1-B data, we identified true race effects in 92%-100% cases and found spurious race effects in at most 16% cases. Both cases revealed relatively accurate identification.

Table 2: Race effects using correct models.

2-B DDM fitted						
Condition	start point		drift NG		drift G	
	$<$	$>$	$<$	$>$	$<$	$>$
2-B: No effect	3	2	2	0	2	6
2-B: drift NG	4	1	0	**48**	2	6
2-B: drift G	5	1	2	0	**50**	0
2-B: start point	**42**	0	4	1	2	4
1-B DDM fitted						
Condition	boundary		drift NG		drift G	
	$<$	$>$	$<$	$>$	$<$	$>$
1-B: No effect	2	2	1	3	5	3
1-B: drift NG	2	2	**50**	0	5	3
1-B: drift G	2	2	1	3	**46**	0
1-B: boundary	**50**	0	0	3	3	3

Table 3: Race effects using wrong models.

2-B DDM fitted						
Condition	start point		drift NG		drift G	
	$<$	$>$	$<$	$>$	$<$	$>$
1-B: No effect	3	0	0	0	3	1
1-B: drift NG	0	5	**49**	0	2	1
1-B: drift G	4	1	0	0	**39**	0
1-B: boundary	0	**50**	0	26	1	10
1-B DDM fitted						
Condition	boundary		drift NG		drift G	
	$<$	$>$	$<$	$>$	$<$	$>$
2-B: No effect	1	1	0	0	5	17
2-B: drift NG	4	0	0	**4**	15	1
2-B: drift G	1	1	0	0	**50**	0
2-B: start point	0	**27**	0	0	10	5

Race effects using wrong models

We were curious how robust our results would be if we analyzed the data for race effects (or any manipulated effect) using the wrong model. Thus, we next examined whether we could still correctly identify race effects in data using wrong models that are inconsistent with the data-generating mechanism. Results are shown in Table 3. Using the 1-B model to fit 2-B data, we identified true race effects in 8%-100% cases, and found spurious effects in at most 44% cases. Using the 2-B model to fit 1-B data, we identified true race effects in 78%-100% cases, and found spurious effects in at most 52% cases. Overall, using the wrong model increases the chance of missing true effects and finding spurious effects. The 1-B DDM is more likely to miss true effects, while the 2-B DDM is more likely to reveal spurious effects.

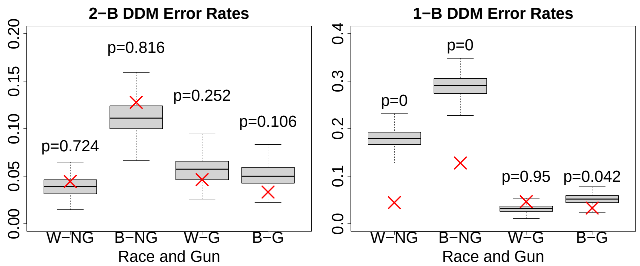

Figure 2: Posterior predictive distributions of error rates from 2-B and 1-B DDMs. The observed error rates are red crosses.

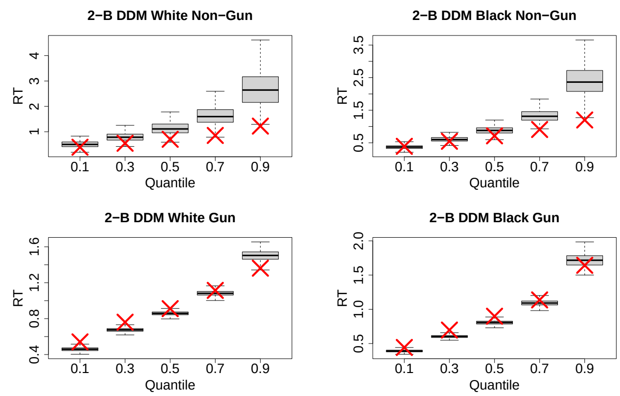

Figure 3: Posterior predictive distributions of RT quantiles from 2-B DDM. Observed quantiles are the red crosses.

Empirical Study

We applied 1-B and 2-B DDMs to an empirical dataset to assess race effects in police academy cadets' decisions to use deadly force. Police use of deadly force has disproportionately impacted Black individuals in United States (Collaborators et al., 2021). To understand how race may enter the decision to shoot, evidence accumulation models have been used as tools for understanding the dynamic aspects of the decision to shoot (Pleskac et al., 2018; Johnson, Stepan, Cesario, & Fenn, 2021; Todd et al., 2021; Johnson, Cesario, & Pleskac, 2018).

In our study, we aimed to achieve two goals. First, we intended to identify the more plausible mechanism (2-B vs 1-B) underlying the decision to shoot. Second, we sought to identify potential race effects in the data and their underlying mechanisms.

Data and study procedures

Newly enrolled cadets in their first week in a police academy ($N = 54$) completed $I = 40$ trials each. Participation was voluntary and monetary compensation was given. Participants were shown different scenarios via a shooting simulator similar to law enforcement training simulators (Ref withheld). In the simulator, realistic policing scenarios are presented via videos projected at life-size. In each trial, participants witnessed a policing scenario in which a suspect either presented a gun and fired at the participant or presented a harmless object. The participants were instructed to make a Shoot response using a modified Glock handgun if they observed a gun. Responses and RTs were recorded. The participants were not informed that suspect race was a study factor.[3]

We applied both 2-B and 1-B DDMs to the data using "rstan." For each model, we obtained two chains with 500 warmups and 5000 samples each, providing a satisfactory effective sample size. The $\hat{R}$ statistics (Gelman, Shirley, et al., 2011) were less than 1.05 for all parameters, indicating good convergence.

Model comparison and posterior predictive distributions

We examined model fit via model comparison (LOO) and the posterior predictive distributions. The LOO of the 1-B DDM is higher than that of the 2-B DDM (Diff=58.4, SE=20.5), indicating that the 2-B DDM was a better fit.

Figure 2 showed the posterior predictive distributions for error rates in each condition (Race$\times$Gun), the true error rates (red crosses), and posterior predictive p-values (Meng, 1994)[4]. The 2-B DDM had an overall good fit to the data, while the 1-B DDM was not a good fit.

Considering that the 1-B DDM did not fit the data, we only examined the posterior predictive fits for RTs in the better-fitting 2-B DDM. Figure 3 shows the posterior predictive boxplots for five RT quantiles (.1, .3, .5, .7, .9) in each condition (Race$\times$Gun), as well as the corresponding true RT quantiles from data (red crosses). We found that the 2-B DDM overestimated the tail RTs in Non-Gun conditions, while underestimated fast RTs in Gun conditions. This may be due to the 4-parameter DDM's lack of ability to characterize inter-trial variability in DDM parameters (Ratcliff & Rouder, 1998). We may be able to characterize fast errors by incorporating inter-trial variability in the starting point, and characterize fast accurate responses by incorporating inter-trial variability in the drift rate (Henrich, Hartmann, Pratz, Voss, & Klauer, 2024).

Race effects in the decision to shoot

Lastly, we examined what the 2-B and 1-B DDMs reveal about race effects during the decision to shoot. Figure 4 displays the the posterior distributions of group-level drift rate, starting point (2-B), and decision boundary (1-B).

Using the 2-B DDM, we did not find a credible difference between Black and White suspects in the drift rates for the

[3] The cadets also completed this task at the conclusion of the academy. We focus on time point 1 in this paper.

[4] A posterior predictive p-value between .05 and .95 indicates reasonable fit.

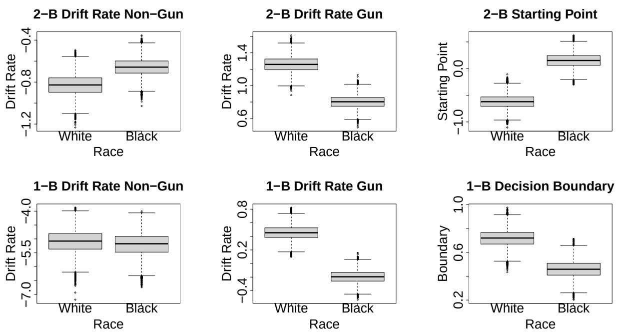

Figure 4: Posterior distributions of group-level parameters.

non-gun trials ($M = 0.17$, 95% HDI (-0.08, 0.44)). When the suspects held a gun, participants had a lower drift rate for Black suspects ($M = -.46$, 95% HDI (-0.71, -0.22)). This result is counter to a stereotypical race effect and to some past results (Pleskac et al., 2018, 2024). However, the cadet participants also had elevated start points for Black vs White suspects ($M = 0.77$, 95% HDI (.44, 1.13)). This result is consistent with a stereotypical race effect and explains the observed elevated error rates for unarmed Black suspects (Figure 2).

The 1-B DDM revealed similar results. There was not a credible race effect in drift rates for the non-gun condition ($M = -0.10$, 95% HDI (-1.34, 1.03)). There was a credible effect for the gun condition with a lower drift rate for the Black suspects ($M = -0.26$, 95% HDI (-0.93, -0.38)). The single decision boundary was lower for Black vs White suspects ($M = -0.26$, 95% HDI (-0.46, -0.06)), indicating cadet participants started closer to Shoot for Black suspects.

In summary, despite having differential model fit, both models provided relatively consistent information about mechanisms underlying race effects for the cadet participants. Neither model revealed a race difference in the evidence accumulation rate when the suspects did not have a gun. However, when the suspects had a gun, both models suggested that the cadets had a faster speed in accumulating evidence towards a Shoot decision when the suspects were White. Meanwhile, the cadets had higher relative start points towards Shoot for Black suspects in the 2-B DDM. A parallel effect appeared for the 1-B DDM where the single boundary was closer to the start point for Black suspects.

Discussion

The Go/No-Go response task is widely used in cognitive science. Closer scrutiny reveals that this response task appears throughout our everyday lives, such as when people decide to cross the street, employees decide to respond to a late-night email, or when police have to make a decision to use deadly force. This study understands the Go/No-Go response processes by investigating two alternative models: 1-B and 2-B DDMs. A simulation study was performed to obtain guidelines for the implementation of these models. The simulation results suggested that Bayesian model comparison methods, such as LOO, may distinguish 1-B and 2-B mechanisms with high precision using Bayesian hierarchical modeling. We also found that the effects in data, such as race differences during the decision to shoot, are more likely to be revealed using a correct model that is consistent with the underlying mechanism. Therefore, in Go/No-Go modeling, we recommend implementing both models, performing a model comparison using LOO to select the better-fitting model, and interpreting results from the better-fitting model. In this way, Go/No-Go studies are more likely to identify true effects in data and less likely to find spurious effects.

We applied the models to an empirical dataset to understand the response processes and race effects during the decision to shoot in a simulator. Model comparison and model fit results endorsed the 2-B DDM over the 1-B DDM. This may indicate that, at least during the simulator, cadets make active No-Shoot decisions and determine suspects to be harmless in policing scenarios. Future work will examine how training and police experience may impact this response process. In terms of race effects, behaviorally, the cadets who had newly arrived at the academy were less accurate for unarmed Black suspects. At the cognitive level, both the 2-B and 1-B models revealed consistent race effects: the cadets accumulate evidence towards the Shoot decision faster when they face an armed White suspect, but the amount of evidence needed to make a Shoot decision for a white suspect is higher than that for a black suspect. Altogether, both models isolated the increased error rate for Black unarmed suspects to a heightened initial proclivity toward the Shoot response for Black suspects. At the same time, the models imply that the similar error rates for armed suspects are due to differences in the start points and drift rates canceling themselves out. Past work has typically isolated any race differences we observe at the behavioral level to evidence accumulation rates (Pleskac et al., 2018; Correll, Wittenbrink, Crawford, & Sadler, 2015; Pleskac et al., 2024). However, there is some evidence for police showing differences in the start points for the much simpler First Person Shooter task (Johnson et al., 2018). Future work will examine how training and experience impact these effects and will use our improved modeling methods to link eye-tracking data to these processes.

This study is not without limitations. Firstly, we made an assumption that all participants share the same underlying mechanism when processing the same type of Go/No-Go stimuli. In reality, it is plausible that different individuals follow differential mechanisms. For example, some police officers may make active No-Shoot decisions and deem suspects as harmless in their policing practices, while some other police officers may stay alert and make no definitive No-Shoot decisions. In future studies, we intend to design and build methods that determine underlying mechanisms based on specific individuals instead of group-level behaviors. We will also explore how model-averaging methods may help us reach conclusions robust to this limitation.

Secondly, while we suggested that participants could have differential Go/No-Go processes based on experimental conditions and stimuli, this study only identified an empirical dataset following the 2-B mechanism. In future studies, we intend to investigate a wider variety of data to identify conditions that result in 1-B processes.

Thirdly, our study is restricted to the condition where there are limited trials per participant, and we placed strong restrictions on the Bayesian priors as a result. However, it is noteworthy that both the selection of Bayesian priors and the trial numbers per participant may affect model fit and model comparison results. Therefore, we intend to expand the simulation to cover a broader range of conditions and obtain more nuanced guidelines for implementing evidence accumulation models for the Go/No-Go paradigm.

Lastly, we adapted the 4-parameter version of DDM for the 2-B DDM, which lacks the ability to characterize inter-trial variability in starting points and drift rates. It may resulted in mis-estimations in the RT distributions. To mitigate this issue, we intend to upgrade the 2-B DDM with the 7-parameter version of DDM (Henrich et al., 2024) in future studies.

Acknowledgment

This study is funded by National Science Foundation Grant No. 1756092 to J.C., T.S., and T.J.P.

References

Ba, Y., Zhang, W., Salvendy, G., Cheng, A. S., & Ventsislavova, P. (2016). Assessments of risky driving: A go/no-go simulator driving task to evaluate risky decision-making and associated behavioral patterns. *Applied ergonomics*, *52*, 265–274.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153–178.

Collaborators, G. . P. V. U. S., et al. (2021). Fatal police violence by race and state in the usa, 1980–2019: a network meta-regression. *The Lancet*, *398*(10307), 1239–1255.

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: how stereotypes disambiguate visual stimuli. *Journal of personality and social psychology*, *108*(2), 219.

Donders, F. C. (1969). On the speed of mental processes. *Acta psychologica*, *30*, 412–431.

Eagle, D. M., Bari, A., & Robbins, T. W. (2008). The neuropsychopharmacology of action inhibition: cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology*, *199*, 439–456.

Gelman, A., Shirley, K., et al. (2011). Inference from simulations and monitoring convergence. *Handbook of markov chain monte carlo*, *6*, 163–174.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389.

Henrich, F., Hartmann, R., Pratz, V., Voss, A., & Klauer, K. C. (2024). The seven-parameter diffusion model: an implementation in stan for bayesian analyses. *Behavior Research Methods*, *56*(4), 3102–3116.

Johnson, D. J., Cesario, J., & Pleskac, T. J. (2018). How prior information and police experience impact decisions to shoot. *Journal of Personality and Social Psychology*, *115*(4), 601–623. doi: 10.1037/pspa0000130

Johnson, D. J., Stepan, M. E., Cesario, J., & Fenn, K. M. (2021). Sleep deprivation and racial bias in the decision to shoot: a diffusion model analysis. *Social Psychological and Personality Science*, *12*(5), 638–647.

Matzke, D., Logan, G. D., & Heathcote, A. (2020). A cautionary note on evidence-accumulation models of response inhibition in the stop-signal paradigm. *Computational Brain & Behavior*, *3*, 269–288.

Mekawi, Y., & Bresin, K. (2015). Is the evidence from racial bias shooting task studies a smoking gun? results from a meta-analysis. *Journal of Experimental Social Psychology*, *61*, 120–130. doi: 10.1016/j.jesp.2015.08.002

Meng, X.-L. (1994). Posterior predictive *p*-values. *The annals of statistics*, *22*(3), 1142–1160.

Meule, A. (2017). Reporting and interpreting task performance in go/no-go affective shifting tasks. *Frontiers in Psychology*, *8*, 701.

Myers, C. E., Dave, C. V., Callahan, M., Chesin, M. S., Keilp, J. G., Beck, K. D., … others (2023). Improving the prospective prediction of a near-term suicide attempt in veterans at risk for suicide, using a go/no-go task. *Psychological Medicine*, *53*(9), 4245–4254.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social cognition*, *19*(6), 625–666.

Pleskac, T. J., Cesario, J., Johnson, D., & Gagnon, G. (2024). Modeling police officers' deadly force decisions in an immersive shooting simulator. *OSF, https://doi.org/10.31234/osf.io/ms23h*.

Pleskac, T. J., Cesario, J., & Johnson, D. J. (2018). How race affects evidence accumulation during the decision to shoot. *Psychonomic Bulletin & Review*, *25*, 1301–1330.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Ratcliff, R., Huang-Pollock, C., & McKoon, G. (2018). Modeling individual differences in the go/no-go task with a diffusion model. *Decision*, *5*(1), 42.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic bulletin & review*, *9*(3), 438–481.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., … others (2001). Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage*, *13*(2), 250–261.

Rubia, K., Smith, A., & Taylor, E. (2007). Performance of children with attention deficit hyperactivity disorder (adhd) on a test battery of impulsiveness. *Child Neuropsychology*, *13*(3), 276–304.

Stan Development Team, C., et al. (2020). Rstan: the r interface to stan. *R package version 2.21. 2*.

Steingroever, H., Wabersich, D., & Wagenmakers, E.-J. (2021). Modeling across-trial variability in the wald drift rate parameter. *Behavior Research Methods*, *53*, 1060–1076.

Todd, A. R., Johnson, D. J., Lassetter, B., Neel, R., Simpson, A. J., & Cesario, J. (2021). Category salience and racial bias in weapon identification: A diffusion modeling approach. *Journal of personality and social psychology*, *120*(3), 672.

Trueblood, J. S., Endres, M. J., Busemeyer, J. R., & Finn, P. R. (2011). Modeling response times in the go/no-go discrimination task. In *Cogsci... annual conference of the cognitive science society. cognitive science society (us). conference* (Vol. 2011, p. 1866).

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, *27*, 1413–1432.

Wang, Z. J., & Busemeyer, J. R. (2021). *Cognitive choice modeling*. MIT Press.

Wright, L., Lipszyc, J., Dupuis, A., Thayapararajah, S. W., & Schachar, R. (2014). Response inhibition and psychopathology: a meta-analysis of go/no-go task performance. *Journal of abnormal psychology*, *123*(2), 429.