# Supplemental Materials for
# Modeling Police Officers' Deadly Force Decisions in an Immersive Shooting Simulator


Timothy J. Pleskac*
University of Indiana, Bloomington

Joseph Cesario*
Michigan State University

David J. Johnson*
Michigan State University

Glen Gagnon
Michigan State University

# Contents

## Scenario Descriptions

We created eight different policing scenarios officers commonly face, based on the expertise of officers from the Milwaukee Police Department. We provide descriptions of the two versions of each scenario. Within each version there were two versions one where the suspect was compliant and one where they were not. We describe each of these versions here.

### Pullover (Day)

The officer is given dispatch information that they have pulled over a driver for a minor traffic infraction.

**Compliant version.** The driver exists the vehicle while the officer is at his vehicle, behind his door. The driver explains that he has to go, and then agrees to get back into his car. He then reaches into the car and pulls out either a cell phone or a gun. The armed driver shoots at the officer, while the unarmed driver explains that he has to go now because his wife called and is having an emergency.

**Noncompliant version.** The driver exists the vehicle while the officer is at his vehicle, behind his door. He is agitated. He yells at the officer that he wasn't speeding. The driver explains that he has to go, and then agrees to get back into his car. He then reaches into the car and pulls out either a cell phone or a gun. The armed driver shoots at the officer, while the unarmed driver explains that he has to go now because his wife called and is having an emergency.

### Pullover (Night)

The officer is given dispatch information that they have pulled over a driver for a minor traffic infraction.

**Compliant version.** The officer walks towards the car and stops by the driver's door. The driver tells the officer that he is going to get his registration and then reaches into his glove box, pulling out either registration or a gun. The armed driver shoots at the officer, while the unarmed driver hands the officer his registration.

**Noncompliant version.** The officer walks towards the car and stops by the driver's door. The driver is agitated and curses at the officer for making him late for work. He reaches into his glove box, pulling out either registration or a gun. The armed driver shoots at the officer, while the unarmed driver hands the officer his registration.

### Parking Lot

The officer is given dispatch information that a suspicious person is looking into cars in a parking lot where cars have been vandalized.

**Compliant version.** The officer sees a suspect looking into cars. The suspect turns and talks to the officer. He explains that he is looking for his wife's rental car and doesn't know which one it is. He reaches into his pocket to show the officer the keys and his identification. He pulls out either his identification or a gun. The armed suspect shoots the officer, while the unarmed suspect holds out his ID for the officer to inspect.

Table S1

*Suspect race and scenario appearance.*

| Actor | Race | Scenario 1 | Scenario 2 |
|---|---|---|---|
| 1 | Black | Parking Lot | Pullover (Night)* |
| 2 | Black | Parking Lot* | Pullover (Night) |
| 3 | Black | Alley* | NA |
| 4 | Black | Alley | NA |
| 5 | Black | Domestic | Sidewalk* |
| 6 | Black | Domestic* | Sidewalk |
| 7 | Black | Pullover (Day) | Warehouse* |
| 8 | Black | Pullover (Day)* | Warehouse |
| 9 | Black | Grocery | NA |
| 10 | Black | Grocery* | NA |
| 11 | White | Parking Lot* | Pullover (Night) |
| 12 | White | Parking Lot | Pullover (Night)* |
| 13 | White | Domestic* | Sidewalk |
| 14 | White | Domestic | Sidewalk* |
| 15 | White | Pullover (Day)* | Warehouse |
| 16 | White | Pullover (Day) | Warehouse* |
| 17 | White | Grocery | NA |
| 18 | White | Grocery* | NA |
| 19 | White | Alley* | NA |
| 20 | White | Alley | NA |

*Note.* Suspect numbers match the numbers used in the main text. *Suspect was “non-compliant.”

**Noncompliant version.** The officer sees a suspect looking into cars. The suspect turns and talks to the officer. He is agitated, and curses at the officer, telling him to leave him alone. After walking away from the officer he turns back to ostensibly show him his identification. He reaches into his pocket and pulls out either his identification or a gun. The armed suspect shoots the officer, while the unarmed suspect holds out his ID for the officer to inspect.

## Domestic

The officer is given dispatch information that a woman wants her boyfriend removed from her house. The ethnicity of the woman always matches that of her boyfriend.

**Compliant version.** The girlfriend explains that her boyfriend is smoking marijuana in her home and has taken her cell phone. The officer follows her into a small sitting room where the boyfriend is on the couch. He asks who the officer is and says that he hasn’t done anything. The boyfriend says that he doesn’t want any trouble and moves to leave. He grabs his backpack and either pulls out a gun or the girlfriend’s cell phone. The armed boyfriend shoots the girlfriend, while the unarmed boyfriend throws the phone at the girlfriend while cursing at her.

**Noncompliant version.** The girlfriend explains that her boyfriend is smoking marijuana in her home and has taken her cell phone. The officer follows her into a small sitting room where the boyfriend is on the couch. He curses at the officer and says that he hasn't done anything. The girlfriend asks where her phone is and the boyfriend yells at her. He grabs his backpack and either pulls out a gun or the girlfriend's cell phone. The armed boyfriend shoots the girlfriend, while the unarmed boyfriend throws the phone at the girlfriend while cursing at her.

### Grocery

The officer is given dispatch information that a person is shoplifting from a grocery store.

**Compliant version.** The officer walks into the store and is greeted by a White male shop owner. He thanks the officer for coming and explains that there is a person shoplifting in the store who did the same thing last week. The officer approaches the suspect, who is putting food into his backpack. The suspect explains that he's just trying to get his money back from last week. The suspect apologizes and drops his backpack. He says he will put the food back as he reaches into his backpack. He either pulls out some food and puts it on the shelf or pulls out a gun and fires at the officer.

**Noncompliant version.** The officer walks into the store and is greeted by a White male shop owner. He thanks the officer for coming and explains that there is a person shoplifting in the store who did the same thing last week. The officer approaches the suspect, who is putting food into his backpack. The suspect is agitated. He explains that the store ripped him off last week, and that he's just trying to get back what he's owed. The suspect curses at the officer and reaches into his backpack. He either pulls out food and explains that the officer can't stop him from taking it, or pulls out a gun and fires at the officer.

### Warehouse

The officer is given dispatch information that a suspicious person has been seen breaking into a building at night.

**Compliant version.** The officer arrives and sees the suspect trying to wedge open the door to a building with a crowbar. The suspect sees the officer and drops the crowbar in surprise. He explains to the officer that he's trying to get out of the cold and doesn't want any trouble. He reaches for his waistband as he crouches to the ground. He either draws a gun out of his waistband and shoots at the officer, or puts his hand up in front of his face after he finishes crouching.

**Noncompliant version.** The officer arrives and sees the suspect trying to wedge open the door to a building with a crowbar. The suspect sees the officer and drops the crowbar. He is agitated and explains to the officer that he's trying to get into his building but has lost his keys. He picks up the crowbar and attempts to pry open the door again. He finally responds to the officer by dropping the crowbar and telling the officer that he'll show him his identification. He reaches into his jacket and either pulls out a wallet or a weapon. The armed suspect shoots the officer, while the unarmed suspect holds out his ID for the officer to inspect.

### Sidewalk

The officer is given dispatch information that a suspect who has committed an armed robbery a few days ago has been positively identified.

**Compliant version.** The officer sees the suspect outside on a sidewalk talking on a phone and walks towards him. The suspect glances around anxiously while waking backwards and says that the officer is arresting the wrong man. As the officer walks closer the suspect says that he is going to put his phone away and puts his hands in his jacket pockets. He pulls his hands out and either pulls out a gun or just his empty hands. The armed suspect shoots the officer, while the unarmed suspect puts his hands up.

**Noncompliant version.** The officer sees the suspect outside on a sidewalk and walks towards him. The suspect is agitated and curses at the officer while putting his hands in his pockets. He pulls his hands out and either pulls out a gun or just his empty hands. The armed suspect shoots the officer, while the unarmed suspect points his middle fingers at the officer.

### Alley

The officer is given dispatch information that a suspect is wanted for arrest for a non-violent crime.

**Compliant version.** The officer approaches the suspect, who is smoking in an alley near a door to a building. He sees the officer and tells him that he has the wrong person and that he didn't do anything. While he is doing this he glances around anxiously. He drops his cigarette and reaches into his back pocket. He either pulls out a wallet or a gun. The armed suspect shoots the officer, while the unarmed suspect holds out his wallet for the officer.

**Noncompliant version.** The officer approaches the suspect, who is smoking in an alley near a door to a building. He sees the officer and curses, running away from the officer. After about twenty feet he turns towards the officer and puts his hand behind his back and yells "Don't Shoot!" He either pulls out a gun or a cell phone. The armed suspect shoots the officer, while the unarmed suspect holds his cell phone.

## Behavioral Analyses

### Modeling scenario and actor variability

Decision errors were analyzed using multilevel logistic regression with suspect race (White, Black), object (gun, nongun), behavior (antagonistic, nonantagonistic), and their interactions as fixed effects. Response times for armed targets were analyzed using multilevel linear regression with suspect race and behavior as a fixed effect. We modeled the variability between participants by including random intercepts for officers. Initial examination of the data revealed sizeable heterogeneity in behavior between scenarios and suspects. To determine how to model the variability best we carried out a model comparison comparing four different methods for modeling variability:

- Random intercepts for scenarios and actor (RI-S,A).
- Random intercepts for each scenario and suspect combination (i.e., each video) (RI-V).

- Random intercepts for scenarios and suspects with random slopes for each manipulated variable nested within the scenario (RI-S,A; RS-C within S).

- Random intercepts for scenarios and suspects with random slopes for each manipulated variable nested within actor (RI-S, C; RS-C within A).

We compared these four models using leave-one-out (LOO) cross-validation for error rates and response times for armed suspects.

Table S2

*Comparison of models of different models of scenario and actor variability for the error rates.*

| | elpd_diff | se_diff | elpd_loo | se_elpd_loo | p_loo | se_p_loo | looic | se_looic |
|---|---|---|---|---|---|---|---|---|
| RI-S,A; RS-C w S | 0 | 0 | -3629.287 | 75.834 | 384.884 | 9.603 | 7258.574 | 151.668 |
| RI-S,A; RS-C w A | -1.805 | 1.74 | -3631.092 | 75.896 | 386.665 | 9.679 | 7262.183 | 151.793 |
| RI -V | -2.337 | 1.776 | -3631.624 | 75.912 | 386.896 | 9.686 | 7263.247 | 151.823 |
| RI-S,A | -441.339 | 30.732 | -4070.626 | 85.771 | 301.165 | 7.042 | 8141.252 | 171.542 |

**Error rate.** Table S2 lists the results from the model comparisons. The best performing model was one with random intercepts for scenarios and suspects and random slopes for the conditions nested within scenarios.

Table S3

*Comparison of models of different models of scenario and actor variability for the response times of armed suspects.*

| | elpd_diff | se_diff | elpd_loo | se_elpd_loo | p_loo | se_p_loo | looic | se_looic |
|---|---|---|---|---|---|---|---|---|
| RI -V | 0 | 0 | 56.004 | 311.128 | 622.541 | 47.294 | -112.007 | 622.256 |
| RI-S,A; RS-C w S | -1.454 | 1.479 | 54.55 | 311.89 | 624.629 | 48.197 | -109.099 | 623.78 |
| RI-S,A; RS-C w A | -3.316 | 1.996 | 52.688 | 312.328 | 628.416 | 49.345 | -105.376 | 624.655 |
| RI-S,A | -901.813 | 58.913 | -845.809 | 268.613 | 585.084 | 39.437 | 1691.618 | 537.226 |

**Response Times.** Table S3 lists the results from the model comparisons. The best-performing model was one with random intercepts for each video, so each unique scenario by actor by scenario combination. The second best model was the model where there were random intercepts for suspects and actors and random slopes for the conditions nested within scenarios. It was within the margin of error with the best model. When we examined the coefficients in terms of the effects of race, suspect behavior, and their interaction, the conclusions from both models were identical. Hence, for consistency with the error rates, in the main paper we report the results from the model with random intercepts for scenarios and suspects with random slopes for each manipulated variable nested within the scenario. Summaries from all four models are provided on the OSF site for this paper.

### Retrospective Power Analysis

The three-way interaction between suspect race, weapon presence, and suspect behavior was unexpected, particularly the higher error rate for Black unarmed nonantogonistic suspects relative to Whites. Besides examining the variability in this effect across scenarios and suspects, we also carried out a retrospective Bayesian power analysis to better appreciate the robustness of our results (J. Kruschke, 2014; J. K. Kruschke & Liddell, 2018; J. K. Kruschke, 2010). In this analysis, we asked, "Given our posterior distribution for the effect size, what is the probability that a new dataset (of the same size and design) will produce the same conclusion?" To do this we, followed the following steps.

- Using the estimated posterior distribution from the logistic regression model predicting error rates, we generated $N = 100$ simulated datasets for the exact same study. In effect, these datasets each correspond to our current beliefs (posterior) about the effect size. Given our current beliefs, what new datasets might we see if we were to repeat the experiment?

- We then fit the logistic regression model to each dataset. To do this, we used the standard priors that we used in the original dataset where the prior for each regression coefficient was set to be normally distributed with means $M = 0$ and standard deviation $SD = 2.5$. Thus, our analysis assumed a skeptical audience starting our analysis not with priors based on our data but with priors set around zero (as if we had not run a study).

- For each simulated dataset, we calculated the mean posterior prediction and credible interval for the three-way interaction. For each statistic, we also asked if the credible interval excluded 0.

Our conclusions proved robust. For the standard priors, the average posterior mean for the three-way interaction between suspect race, weapon presence, and suspect behavior was $M = 2.36$ ($SD = 0.57$). The average lower and upper points for the credible interval were $M = 0.55$ ($SD = 0.57$) and $M = 4.23$ ($SD = 0.65$), respectively. Though slightly higher, these estimated values are within the same range as our estimated values of $b = 2.21$, [0.37, 4.09]. Across all 100 simulated datasets, the credible intervals excluded zero 81 out of 100 (81%) of the time. This analysis, thus, suggests a new study with the same design has a good chance of replicating our results.

### Additional behavioral analyses

Beyond whether officers fired and the response times associated with the first shot (discussed in the main text), we examined three additional behavior-level outcomes: 1) the number of shots fired, 2) whether an officer grabbed their weapon (as defined by fully wrapping their hand around the grip of the pistol), and 3) the response time associated with grabbing the weapon. We analyzed these data with multilevel regression using the rstanarm package (Goodrich et al., 2018) in R. We controlled for variability across participants and stimuli by allowing intercepts to vary across officers, scenarios, and suspects (Judd et al., 2012). Consistent with the error rate and RT models, we also set the models to have random slopes for each manipulated variable nested within the scenario. We used the default priors in rstanarm (version 2.18.2), which are weakly informative. We report the posterior predicted mean of the parameter or statistic of interest and (in brackets) its 95% Highest Density Interval (HDI; J. Kruschke, 2014). The HDI summarizes the posterior distribution such that values within the 95% HDI indicate the most credible values. Thus, we use the term *credible* throughout.

The decision to grab one's weapon was analyzed with logistic regression, while the number of shots fired was examined with Poisson regression. Effects coding was used to indicate suspect race (White = -.5, Black = .5) and object (nongun = -.5, gun = .5). To be comparable to previous research, we anchored the response time associated with the first shot fired based on when the object in the citizen's hand was first visible. The response time to grab one's weapon was measured from the start of the scenario.

One limitation of our experimental design was that each officer was exposed to the same scenario multiple times. This may have led to learning effects and artifacts in the data. To avoid these issues, we examined a subset of the data ($N$ = 5,117, 26% of trials) that included only the officers' first exposure to a given scenario. Our results remain unchanged when examining this subset of data.

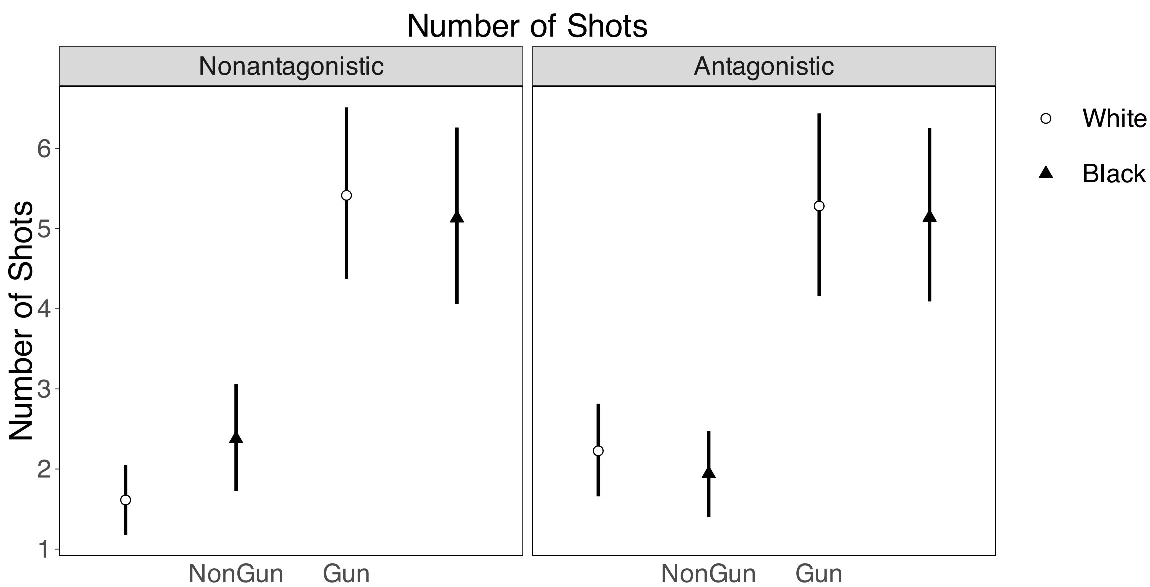


*Figure S1*. Number of shots fired by race. Dots indicate mean posterior estimate and bars indicate 95% HDI.

**Shots Fired.** Figure S1 shows the number of shots fired on any trial that officers fired at least once ($N$ = 10,286, 53% of total trials). Officers shot armed suspects ($M$ = 5.2 [4.4, 6.1]) more than unarmed suspects ($M$ = 2.0 [1.6, 2.4]), $b$ = 0.96 [0.78, 1.17]. Overall, the number of shots fired did not credibly vary based on suspect race, $b$ = -0.04 [-0.18, 0.10]. However, there was a credible three-way interaction between suspect race, weapon presence, and suspect behavior $b$ = 0.552, [0.152, 0.947]. This interaction was driven by the effect of race on nonantagonistic unarmed suspects. More shots were fired for unarmed nonantagonistic Black suspects ($M$ = 2.38 [1.72, 3.07] than unarmed nonantagonistic White suspects ($M$ = 1.62 [1.19, 2.04]; $M$ = 0.39, [0.07, 0.68]). These results align with the error rates reported in the main paper where officers had a higher error rate for unarmed, nonantagonistic Black suspects compared to unarmed nonantagonistic White suspects.

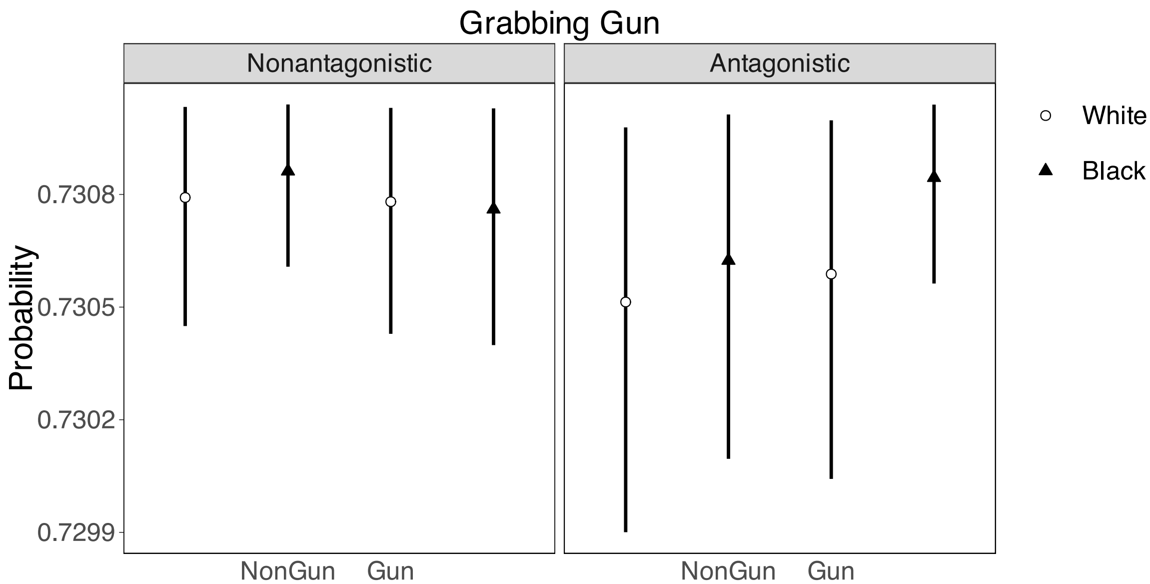


*Figure S2*. Probability of officer grabbing the gun. Dots indicate mean posterior estimate and bars indicate 95% HDI.

**Grabbing Gun.** Were officers more likely to grab their guns for Black relative to White suspects? Figure S2 plots the likelihood of officers grabbing their guns during the first eight trials of the simulator ($N$ = 3,656). Officers uniformly grabbed their weapons for all suspects ($M$ = .73 [.73, .73]). There were no credible effects of object, race, suspect behavior, or any interaction.

### Grabbing Gun Response Time

Did officers grab their gun sooner for Black versus White suspects? Figure S3 plots when officers grabbed their gun relative to the start of the scenario ($N$ = 3,656). Officers uniformally reached for their weapon at 13.8 s [9.6, 16.7] after the scenario started. There were no credible effects of object, race, suspect behavior, or any interaction.

### Variation Across Officer, Scenario, and Suspect

One question that is difficult to assess with other shooting simulators is just how much officers vary from one another, and how this compares to variability from other sources such as the scenario officers face or suspect behavior. To address this question, Table S4 lists intra-class correlations for each outcome across officers, policing scenarios, and suspects. The intra-class correlation quantifies how related the measurement of the outcome variable is to each member of the particular group. Consistent with officer accounts that the main driver of police use of force is the specifics of the situation (Loughlin & Flora, 2017; Rylant, 2017), variability in the decision to shoot was primarily associated with the scenario and suspect. Variability associated with officers was much smaller.

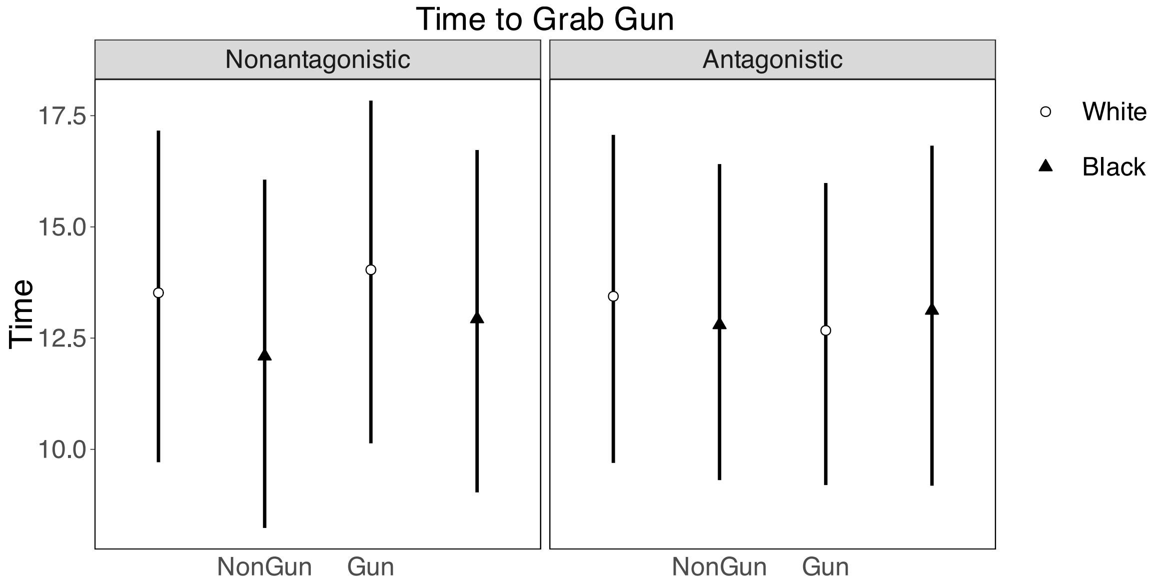


*Figure S3*. Time for officer to grab the gun relative to the start of the scenario. Dots indicate mean posterior estimate and bars indicate 95% HDI.

Table S4
*Intra-class correlations for outcome variables*

| | | Grouping Variable | | |
|---|---|---|---|---|
| Outcome Variable | Scale | Officer | Scenario | Suspect |
| Shooting Decisions | Logistic | .07 [.04, .09] | .19 [.05, .38] | .12 [.05, .22] |
| Shooting Response Time | Identity | .19 [.15, .24] | .14 [.03, .30] | .11 [.05, .19] |
| Number of Shots Fired | Log | .06 [.05, .06] | .03 [.00, .07] | .01 [.00, .01] |
| Gun Grabbed | Logistic | .71 [.61, .81] | .02 [.00, .06] | .00 [.00, .01] |
| Gun Grabbed Response Time | Identity | .15 [.08, .22] | .47 [.24, .71] | .02 [.00, .04] |

*Note.* Brackets indicate 95% HDI.

However, Table S4 also shows that in terms of the other variables, variability was primarily associated with officer and scenario differences rather than suspect differences. For example, for shooting response times, variability was primarily associated with officer and scenario effects rather than suspect effects. The number of shots fired, decision to grab the gun, and the response time to grab followed a similar pattern. Variability was primarily associated with officer and scenario effects rather than suspect effects. (The large scenario ICC for the response time in grabbing a gun is due to differences in scenario length.) Taken together, this suggests there is variability in how much the scenario or suspect behavior is associated with differences across the various types of force, but also there is substantial heterogeneity among officers in their use force in this immersive task.

## Go Diffusion Decision Model

To examine the cognitive processes underlying officers' deadly force decisions, we adapted the Bayesian Diffusion Decision Model framework used for the first-person shooter task

(Johnson et al., 2018; Pleskac et al., 2018) for the immersive shooting simulator. According to the DDM, when deciding whether to shoot or not, participants begin with an initial bias toward one decision or the other and accumulate evidence supporting each option by repeatedly sampling information from the environment relevant to this decision. See Table 1 for a description of the model parameters and Figure 1 for a model realization.

In the first-person shooter task, respondents must press a button indicating a "Shoot" response and a separate button to indicate a "Don't Shoot" response. In this case, according to the model, there are two thresholds — one for each choice. When the evidence reaches either threshold the corresponding decision is made. However, in the immersive shooting simulator (ISS) officers are asked to make a shoot response by pulling the trigger of a handgun. This means that there is only one observed response, a "Shoot" response. Psychologists refer to this type of task as a go/no-go procedure; participants must respond to one option but withhold their response to another (Donders, 1868/1969). The different response type, however, does not necessarily imply a different process. Extensive modeling work has shown that people in a go/no-go procedure use the same evidence accumulation process as they do when they complete a two-alternative forced-choice task (Gomez et al., 2007; Ratcliff et al., 2018). The difference is that the no-go response (i.e., "Don't Shoot") is implicit (Gomez et al., 2007). That is, participants do decide not to shoot at some point, but its timing is not explicitly known. In our Bayesian framework, we implemented this implicit boundary for the "Don't Shoot" response by treating the response time for this response as missing data. This was done by explicitly modeling the probability that the response time was missing, which in this case was the probability of a "Don't Shoot" response (J. Kruschke, 2014). Model recovery analyses confirmed (see below) that the model parameters can be accurately recovered for this adaptation of the DDM, the GoDDM, with the experimental design used in this study.

Past models of a Go/No Go task with a DDM either modeled the response times of the go trials alone (Trueblood et al., 2011) or modeled the response times of the go trials and the accuracy rate(Gomez et al., 2007). Our approach is analogous to the latter and allows us to implement the model in a Bayesian hierarchical manner. Such a framework is potentially applicable to a wider range of Go/No-Go tasks (Chen et al., 2025).

Another important aspect of the ISS is that participants watched and responded to a video that lasted on average 25 seconds, during which different types of information were revealed at different stages. This setup lends itself to modeling the evidence accumulation process via multiple processing stages (Diederich & Busemeyer, 2015; Diederich & Oswald, 2016). However, given the relatively small number of trials per participant and the variability between different scenarios regarding when and what information is revealed, we instead focused on modeling the decision process from when the object (gun or non-gun) was drawn. We used a model comparison to ask how the specific scenario and suspect in the video impacted the model parameters. The scenario and suspect account for the differences between trials regarding the information revealed before the object draw. As we will show, the model comparison revealed that the best model allowed each scenario and suspect to have a different start point, capturing the different information presented before the object was drawn.

### Model Estimation

We used hierarchical Bayesian techniques to estimate the model parameters (J. Kruschke, 2014; Lee & Wagenmakers, 2014; Vandekerckhove et al., 2011). This approach simultaneously models individual and group level differences and allows more precise estimates at the individual level as the data from each participant inform the parameter estimates of other participants.

We also expanded the hierarchical GoDDM to account for potential variability between suspects and scenarios in the videos. Because there are many ways this variability might impact different parameters of the GoDDM, we used a data-driven model comparison approach to specify a model that balanced parsimony and explanatory power (see below). The best-fitting model was one where we allowed variability in the start point across scenarios and suspects. This model can answer questions such as whether initial biases, at least from the point the object is drawn, depend on the suspect and the scenario.

We estimated the model using a Markov Chain Monte Carlo (MCMC) sampler JAGS (Plummer, 2003) with the Wiener distribution provided by Wabersich and Vandekerckhove (Wabersich & Vandekerckhove, 2014). JAGS was run via matjags (Steyvers, 2011) in MATLAB 2018b (via a Windows machine) and all post-processing was done with MATLAB. JAGS and post-processing code are on the OSF website for this paper.

We took an estimation approach to make inferences in this framework (Gelman et al., 2013; J. Kruschke, 2014). We report both the most credible value and the 95% Highest Density Interval (HDI) in brackets to describe the posterior distribution around a parameter of interest. This informed the number of samples we collected via the MCMC sampler. Each reported estimate in the paper was based on an estimated sample size of at least 10,000 after accounting for autocorrelation. Due to high autocorrelation among some parameters, we collected 200,000 samples using 40 chains with 5,000 samples each, a burn-in of 2,000, and recording only every fifth sample. To test differences between conditions, we compute the difference between the conditions and test whether the 95% HDI contains a null value of 0. We conclude that the difference between conditions is credible if it does not.

### Model Recovery

We sought to establish how well we could recover the parameters of the hierarchical GoDDM and also test how well the model could identify a race effect in the relative start point or the drift rates. To do so, we ran two different model recovery analyses: one where the effect of race was in the relative start point for one specific compliance condition (i.e., noncompliance) and one where the effect of race was in the drift rates for one specific compliance condition (i.e., noncompliance). In both cases, we coded the effect to be about 0.75 in terms of standardized mean difference (in terms of between-participant standard deviation). In each simulation, we simulated 50 datasets where $N = 659$ simulated subjects completed 32 trials in an experiment where race, object, and compliance, were orthogonally manipulated with 4 trials in each condition. We used typical values of the hierarchical Bayesian DDM for the First Person Shooter Task (Pleskac et al., 2018) to simulate choice and response times via the stochastic differential equation for the diffusion process (Karlin & Taylor, 1981; Tuerlinckx et al., 2001). The values of the parameters are listed in Table S5. As we sought to establish how well the GoDDM recovered the parameters of a Go/No Go procedure we removed all the response times for the "Don't Shoot" response and replaced them with NaN. Then we estimated the parameters of the GoDDM within the Bayesian hierarchical framework.

Table S5 lists the average posterior mean estimates and the average values of the 95% HDI for each of the generating parameters. The model does a good job recovering the parameters in this experimental design. Due to the smaller sample size per condition, the recovery analyses suggest that there is a small underestimation of the relative startpoint ($\sim .03$) as well as a small overestimation of the threshold separation ($\sim .05$). Moreover, the recovery of the between-participant variability parameter for the relative start point was poorly recovered. Part of the in-precision, particularly in the threshold separation, appears to arise from using random-walk approximation of the diffusion process to simulate the models, But, the other part of the in-precision is due to the small samples per condition at the individual level.

Table S5 also shows that despite some imprecision in parameter estimation, the model can correctly identify whether the race effect was in the relative start point or in the drift rates. This means the sign of the difference is maintained and the magnitude of the difference is maintained (i.e., a low Type S and Type M error rate) (Gelman & Tuerlinckx, 2000; Gelman & Carlin, 2014). Thus, even if a credible effect may not be found the values of the parameters carry information speaking to the effect.

Table S5

*Summary of the parameter recovery analysis for two versions of the GoDDM*

| Description | Parameter | Race Effect in Relative Start Point: Generating Value | Race Effect in Relative Start Point: M 95% HDI | Race Effect in Drift Rate: Generating Value | Race Effect in Drift Rate: M 95% HDI |
|---|---|---|---|---|---|
| Group level relative start point, White, Nonantagonistic | $\mu^{\beta}_{W,NA}$ | 0.529 | 0.497 [0.463, 0.531] | 0.556 | 0.517 [0.481, 0.553] |
| Group level relative start point, Black, Nonantagonistic | $\mu^{\beta}_{B,NA}$ | 0.586 | 0.551 [0.52, 0.583] | 0.556 | 0.53 [0.499, 0.56] |
| Group level relative start point, White, Antagonistic | $\mu^{\beta}_{W,A}$ | 0.556 | 0.521 [0.488, 0.554] | 0.556 | 0.525 [0.493, 0.558] |
| Group level relative start point, Black, Antagonistic | $\mu^{\beta}_{B,A}$ | 0.556 | 0.522 [0.489, 0.555] | 0.556 | 0.523 [0.49, 0.556] |
| Relative Start Point Standard Deviation | $\sigma^{\beta}$ | 0.056 | 0.105 [0.094, 0.116] | 0.056 | 0.106 [0.095, 0.117] |
| Group level threshold separation, White, Nonantagonistic | $\mu^{\alpha}_{W,NA}$ | 1.130 | 1.195 [1.158, 1.233] | 1.130 | 1.185 [1.149, 1.221] |
| Group level threshold separation, Black, Nonantagonistic | $\mu^{\alpha}_{B,NA}$ | 1.130 | 1.188 [1.15, 1.228] | 1.130 | 1.197 [1.158, 1.237] |
| Group level threshold separation, White, Antagonistic | $\mu^{\alpha}_{W,A}$ | 1.130 | 1.188 [1.15, 1.225] | 1.130 | 1.191 [1.153, 1.229] |
| Group level threshold separation, Black, Antagonistic | $\mu^{\alpha}_{B,A}$ | 1.130 | 1.188 [1.15, 1.225] | 1.130 | 1.19 [1.152, 1.227] |
| Threshold separation Standard Deviation | $\sigma^{\alpha}$ | 0.135 | 0.15 [0.12, 0.179] | 0.135 | 0.15 [0.12, 0.178] |
| Group level drift rate Nongun, White, Nonantagonistic | $\mu^{\delta}_{NG,W,NA}$ | -2.330 | -2.208 [-2.425, -1.994] | -2.734 | -2.618 [-2.869, -2.371] |
| Group level drift rate Gun, White, Nonantagonistic | $\mu^{\delta}_{G,W,NA}$ | 2.235 | 2.382 [2.198, 2.566] | 1.831 | 1.975 [1.791, 2.16] |
| Group level drift rate Nongun, Black, Nonantagonistic | $\mu^{\delta}_{NG,B,NA}$ | -2.330 | -2.206 [-2.405, -2.007] | -1.926 | -1.823 [-1.998, -1.648] |
| Group level drift rate Gun, Black, Nonantagonistic | $\mu^{\delta}_{G,B,NA}$ | 2.235 | 2.386 [2.212, 2.56] | 2.639 | 2.736 [2.561, 2.911] |
| Group level drift rate Nongun, White, Antagonistic | $\mu^{\delta}_{NG,W,A}$ | -2.330 | -2.195 [-2.402, -1.99] | -2.330 | -2.228 [-2.435, -2.022] |
| Group level drift rate Gun, White, Antagonistic | $\mu^{\delta}_{G,W,A}$ | 2.235 | 2.399 [2.22, 2.579] | 2.235 | 2.345 [2.168, 2.521] |
| Group level drift rate Nongun, Black, Antagonistic | $\mu^{\delta}_{NG,B,A}$ | -2.330 | -2.189 [-2.394, -1.985] | -2.330 | -2.22 [-2.428, -2.013] |
| Group level drift rate Gun, Black, Antagonistic | $\mu^{\delta}_{G,B,A}$ | 2.235 | 2.386 [2.207, 2.564] | 2.235 | 2.353 [2.176, 2.532] |
| Drift rate standard deviation | $\sigma^{\delta}$ | 0.762 | 0.639 [0.558, 0.719] | 0.762 | 0.623 [0.542, 0.703] |
| Group level relative nondecision time Nongun, White, Nonantagonistic | $\mu^{NDT'}_{NG,W,NA}$ | 0.388 | 0.393 [0.373, 0.412] | 0.388 | 0.393 [0.373, 0.413] |
| Group level relative nondecision time Gun, White, Nonantagonistic | $\mu^{NDT'}_{G,W,NA}$ | 0.360 | 0.354 [0.345, 0.363] | 0.360 | 0.354 [0.344, 0.364] |
| Group level relative nondecision time Nongun, Black, Nonantagonistic | $\mu^{NDT'}_{NG,B,NA}$ | 0.388 | 0.391 [0.375, 0.407] | 0.388 | 0.391 [0.375, 0.407] |
| Group level relative nondecision time Gun, Black, Nonantagonistic | $\mu^{NDT'}_{G,B,NA}$ | 0.360 | 0.354 [0.346, 0.362] | 0.360 | 0.355 [0.348, 0.363] |
| Group level relative nondecision time Nongun, White, Antagonistic | $\mu^{NDT'}_{NG,W,A}$ | 0.388 | 0.392 [0.374, 0.409] | 0.388 | 0.392 [0.374, 0.409] |
| Group level relative nondecision time Gun, White, Antagonistic | $\mu^{NDT'}_{G,W,A}$ | 0.360 | 0.354 [0.345, 0.362] | 0.360 | 0.355 [0.346, 0.363] |
| Group level relative nondecision time Nongun, Black, Antagonistic | $\mu^{NDT'}_{NG,B,A}$ | 0.388 | 0.392 [0.374, 0.409] | 0.388 | 0.391 [0.373, 0.409] |
| Group level relative nondecision time Gun, Black, Antagonistic | $\mu^{NDT'}_{G,B,A}$ | 0.360 | 0.354 [0.345, 0.362] | 0.360 | 0.355 [0.346, 0.363] |
| Non-decision time standard deviation | $\sigma^{NDT'}$ | 0.058 | 0.054 [0.051, 0.056] | 0.058 | 0.054 [0.051, 0.056] |

*Note.* The mean and the 95% HDI values are averaged across the 50 different simulated datasets for each parameter recovery analysis. The model estimates slightly overestimate the threshold separation. This overestimation is an artifact from how the data were simulated. Specifically, the diffusion process was simulated with a stochastic differential equation (Tuerlinckx et al., 2001). With this simulation approach, the time point of passing the threshold cannot precisely be determined from a discrete simulation of a continuous process and as a result the simulated response time tends to be on average a bit longer than observed. NDT = Non-decision Time.

## Model Comparison

The intraclass correlations (Table S4) imply a sizeable amount of the variability in the decision to shoot is attributable to the scenario and the suspect in each scene. Model recovery analyses showed it was impossible to incorporate scenario and suspect effects in each of the four process parameters and accurately recover the values. Therefore, we carried out a model comparison to ask in what process parameter the effect of scenario and suspect should be isolated. We started with base model that allowed a race effect in each parameter and participant level variability in each parameter. This specification has proven quite useful in accounting for responses in several different datasets (Johnson et al., 2018; Pleskac et al., 2018). Then we fit different models where each parameter was allowed to vary also as a function of the scenario or the suspect in the scene. We compared the models with the Deviance Information Criterion (DIC) as measure of goodness of fit where smaller values are better (Spiegelhalter et al., 2002). This information criterion metric is useful for hierarchical models and penalizes complexity models.

The model comparison results are summarized in Table S6. In all cases, a model with the effect of scenario and/or suspect accounted for does better than the base model. In the case of the scenarios, the best model, according to the DIC, placed the effect of the scenario in the location of the relative start point. Similarly, for suspects, according to the DIC the best model located the effect of suspects in the start point. Both of these models are sensible in that the suspect and scenario were present in each video before the suspect drew the object, and thus, according to these results, biased the officers in terms of their initial evidence to shoot or not.

The model's fit with both suspect and scenario effects in the relative startpoint is listed in Table S6. It provides a good fit, though notably its fit is worse than the model with only suspect effects (no scenario effects). We chose to use the full model with variability in the scenario in the relative start point for both the scenario and suspects as the measurement model in the analysis below. We made this decision to (a) align the analyses with the behavioral analyses and (b) also to reflect our *a priori* belief that scenario contributes to this variability.

Future work could model these biases in terms of separate stages of evidence accumulation that ultimately lead to different locations of evidence to shoot or not during each stage of processing (Diederich & Oswald, 2016). But, for the current project, these different stages of processing are accounted for via different relative start point biases present when the suspect draws the object.

## Posterior Predictive Checks

Figure S4 plots the observed proportion of trials an officer shot in the immersive simulator. The figure also shows the posterior predicted distribution from the GoDDM. The model does well in capturing the observed proportion of times the officers shot armed targets though there is variability between officers, which the model misses and is likely due to the small number of trials from each officer. For unarmed suspects the model hedges more towards the modal proportions. The posterior distributions align with the predictions from a logistic regression, though the width of the HDI's are a bit smaller. This difference is because the GoDDM is simultaneously modeling choice and response times so the logistic

Table S6

*DIC for models with different effects of scenario and suspect on the process parameters*

| Model | Location of Effect | DIC |
|---|---|---|
| Base Model | | 7,372 |
| Base Model with Scenario Effects | $\beta$ | 5,473 |
| | $\alpha$ | 5,478 |
| | $\delta$ | 6,043 |
| | NDT | 6,231 |
| Base Model with Suspect Effects | $\beta$ | 3,970 |
| | $\alpha$ | 4,187 |
| | $\delta$ | 4,599 |
| | NDT | 5,894 |
| Base Model with Scenario & Suspect Effects | $\beta$ | 4,575 |

*Note.* DIC = Deviance Information Criterion.

regression is based on a subset of the data the GoDDM is using.

Figure S5 plots for each condition the .1, .3, .5, .7, and .9 quantile of the response time distributions across participants and across trials. Also plotted are the posterior predicted distributions of each of those statistics from the GoDDM. Note we do not plot the quantiles for the individual officers as with such few trials those statistics are not reliable. The figure shows that the GoDDM does reasonably well in capturing the observed quantiles for trials in which the suspect was armed though the model tends to predict a slightly slower response time. When an individual was unarmed the model tends to predict a slower response time, which is especially prominent in the tails of the distribution (i.e., the .7 and the .9 quantiles). Part of the issue is likely due to the small number of trials (32 or fewer) each officer completed combined with the high accuracy rate. This makes estimating and fitting the tails of the response time distributions in the unarmed conditions difficult as there are very few observations in this condition already. We explored ways to improve the fit, including adding trial-by-trial variability in the start points and the drift rates, but this did not improve fits and resulted in less reliable group level parameter estimates.

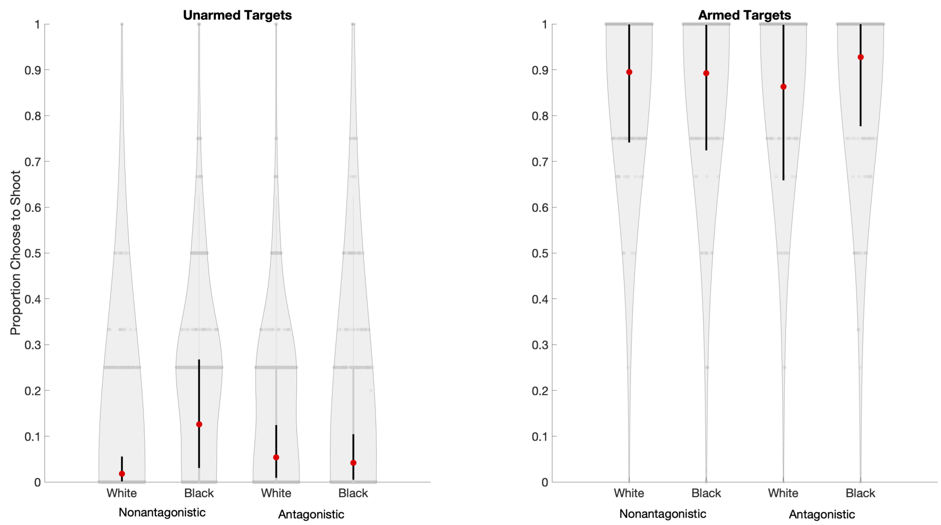


*Figure S4*. Observed and posterior predicted proportion of trials an officer shot in each condition. The observed average proportion are the ×, the small grey circles are the observed proportion for each officer with a distribution estimated over top those dots in light grey, the large red circles are the mean posterior predicted proportion, and the bars denote the 95% HDIs for the posterior predicted distribution across participants and trials. Note the posterior predicted distributions have been adjusted to account for the number of trials each officer completed.

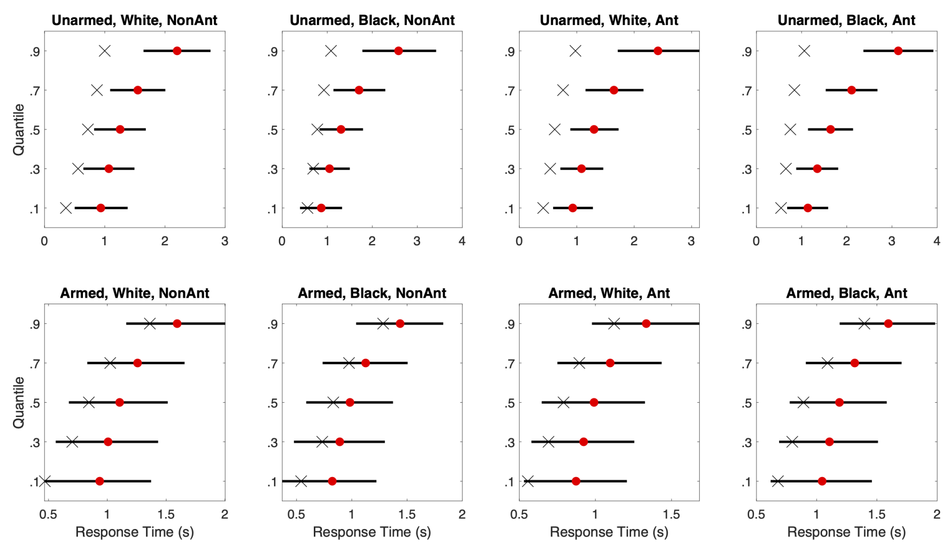


*Figure S5*. Observed and posterior predicted quantiles of the response times in each condition. The observed quantiles are the Xs, the large red circles are the mean posterior predicted quantile, and the bars denote the 95% HDIs for the posterior predicted distribution.

References

Chen, Y., Fang, J., Cesario, J., Liu, T., & Pleskac, T. J. (2025). Comparing one-boundary and two-boundary evidence accumulation models for go/no-go processes: An application to the decision to shoot. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47). doi: 10.31234/osf.io/96nxd_v2

Diederich, A., & Busemeyer, J. R. (2015). *Multi-stage sequential sampling model of multi-attribute decision making.* Jacobs University Bremen, Germany. Retrieved from `https://www.researchgate.net/profile/Adele_Diederich/publication/290434834_Multi-Stage_Sequential_Sampling_Model_of_Multi-Attribute_Decision_Making/links/5697d27108ae34f3cf1f1f09.pdf` (Working paper)

Diederich, A., & Oswald, P. (2016). Multi-stage sequential sampling models with finite or infinite time horizon and variable boundaries. *Journal of Mathematical Psychology*, *74*, 128–145. doi: 10.1016/j.jmp.2016.02.010

Donders, F. C. (1868/1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. doi: 10.1016/0001-6918(69)90065-1

Gelman, A., & Carlin, J. (2014). Beyond power calculations assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. doi: 10.1177/1745691614551642

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis.* Chapman and Hall/CRC.

Gelman, A., & Tuerlinckx, F. (2000). Type s error rates for classical and bayesian single and multiple comparison procedures. *Computational Statistics*, *15*(3), 373–390. doi: 10.1007/s001800000040

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389–413. doi: 10.1037/0096-3445.136.3.389

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). *rstanarm: Bayesian applied regression modeling via Stan.* Retrieved from `http://mc-stan.org/` (R package version 2.17.4)

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

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. doi: 10.1037/a0028347

Karlin, S., & Taylor, H. E. (1981). *A second course in stochastic processes.* Elsevier.

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, jags, and stan.* Academic Press.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300. doi: 10.1016/j.tics.2010.05.001

Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206. doi: 10.3758/s13423-016-1221-4

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.

Loughlin, J. K., & Flora, K. (2017). *Shots fired: The misunderstandings, misconceptions, and myths about police shootings.* Simon and Schuster.

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. doi: 10.3758/s13423-017-1369-6

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.*

Ratcliff, R., Huang-Pollock, C., & McKoon, G. (2018). Modeling individual differences in the go/no-go task with a diffusion model. *Decision*, *5*(1), 42–62. doi: 10.1037/dec0000065

Rylant, C. J. (2017). *Shots fired: The psychology behind officer involved shootings.* Perfect Life.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. doi: 10.1111/1467-9868.00353

Steyvers, M. (2011). *Matjags 1.3: A matlab interface for jags.*

Trueblood, J. S., Endres, M. J., Busemeyer, J. R., & Finn, P. R. (2011). Modeling response times in the go/no-go discrimination task. In *Proceedings of the 33rd annual conference of the cognitive science society* (Vol. 2011, pp. 1866–1871).

Tuerlinckx, F., Maris, E., Ratcliff, R., & De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 443–456. doi: 10.3758/BF03195402

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*(1), 44–62. doi: 10.1037/a0021765

Wabersich, D., & Vandekerckhove, J. (2014). Extending jags: A tutorial on adding custom distributions to jags (with a diffusion model example). *Behavior Research Methods*, *46*(1), 15–28. doi: 10.3758/s13428-013-0369-3