Modeling police officers' deadly force decisions in an immersive shooting simulator

Timothy J. Pleskac[1◑*], Joseph Cesario[2◑*], David J. Johnson[2◑], Glen Gagnon[2]

1 Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, USA
2 Department of Psychology, Michigan State University, East Lansing, MI, USA

◑These authors contributed equally to this work.
*pleskac@iu.edu or cesario@msu.edu

Abstract

We introduce a novel framework to understand how race, suspect behavior, and policing scenario impact police officers' decision to shoot. We report four principle results with a sample of police officers from the Milwaukee Police Department (N=659) that illustrate the utility of this framework: (1) policing scenario and suspect behavior played important roles in officers' decisions; (2) the effect of race on shooting errors depended on whether suspects behaved in an antagonistic or nonantagonistic way; (3) cognitive modeling showed this effect of race was not due to initial biases to shoot Black suspects but instead due to differences in how evidence was gathered between Black and White suspects; and (4) no credible effects of race were observed on response times. Exploration of the data suggests that the race effect may, in part, be due to behaviors performed by particular suspects in specific scenarios. This work provides a novel method and analytic approach for understanding how officers integrate multiple pieces of information during the decision to shoot and how these different sources of information can impact the decision in different ways at different stages. We emphasize that the current report cannot answer the broad question of "Are police in general biased?," but instead is a means to study how officers make deadly force decisions in specific policing scenarios. This sets the stage for researchers and practitioners to obtain the data necessary for designing effective training interventions.

Introduction

Police use of deadly force remains a pressing topic in the U.S. As a result, the question of how officers make the decision to shoot and how factors such as race impact this decision has been of immense interest to researchers across several disciplines. Two main approaches have been taken to understand a police officer's decision to shoot: a laboratory-based approach where officers make a series of decisions in a controlled laboratory setting and an archival-based approach that analyzes actual officer-involved shootings. Each approach brings distinct advantages as well as disadvantages.

The laboratory-based approach comes mainly from psychology and uses controlled laboratory tasks to test for racial bias in the decision to shoot (1; 2; 3). In these tasks, participants typically see static images on a computer screen of Black or White men holding guns or harmless objects. If the suspect is holding a gun, participants are instructed to press a button labeled "Shoot;" otherwise, they press a button labeled

"Don't Shoot." Participants make these decisions across many trials. The precision and control of these tasks allow researchers to study how race impacts the underlying decision process, with the typical explanation that automatic activation of the Black-violent stereotype makes people more likely or faster to "Shoot." The most consistent result is that civilian participants (undergraduate students or untrained community members) "Shoot" the armed Black suspects faster than White suspects and "Don't Shoot" unarmed Black suspects slower than White suspects (2; 4). Some studies also report that under time pressure participants choose "Shoot" for the Black suspects at a greater rate than for White suspects (4; 5; 6; 7; 8), but this effect appears less reliable across studies (2). Trained officers have been found to only exhibit response time differences or show no racial differences at the behavioral level in the decision to shoot (9; 10; 11; 12; 13; 14).

A limitation of this line of research may be the framing of the question itself. By solely focusing on whether officers are more likely to shoot unarmed Black men, the experimental task has centered on manipulating race. But this focus has come at the expense of simplifying or removing factors that officers use when deciding to shoot, such as information about the situation and their interaction with the suspect (1; 10). Removing this information from experimental tasks may constrain the generalizability of conclusions from these studies to actual use-of-force decisions, a limitation that is particularly concerning when it comes to the issue of whether race plays a role in deciding to shoot.

One alternative to the experimental approach is to study deadly force decisions by analyzing actual officer-involved shootings. This archival approach has been primarily taken by researchers in criminal justice. In contrast to the experimental approach, this approach has asked broader questions about the impact of multiple factors on officers' decisions, including the nature of the policing scenario and a suspect's behavior. Such work has shown that (1) situational factors such as the risk level of the encounter or the aggressiveness of the suspect are strongly associated with officers' decisions to shoot and (2) racial disparities in being fatally shot decrease once these factors are taken into account (15; 16; 17; 18; 19; 20; 21; 22; 23; 24; 25; 26).

The archival-based approach is not without its limitations as well. First, this approach offers no possibility for directly studying the cognitive dynamics underlying a given officer's decision to use deadly force, as the lack of experimental control and infrequency of shootings preclude such questions from being asked. Second, although this approach does lend itself to quantifying the evidence for a racial effect once different factors are controlled for (such as the context or civilian behavior, demographic characteristics, or criminal activity, e.g., 15; 17; 23), the uniqueness of each shooting situation raises problems. Factors present in one situation may take on a different meaning in another situation, making ostensibly similar events qualitatively different. Thus, even showing that Blacks are more likely to be shot than Whites in a given scenario (e.g., serving a warrant) makes it difficult to draw general conclusions about racial differences. Coupled with significant county-to-county variation in the decision to shoot (27) and the infrequency with which officers fire their weapons, this approach imposes limitations on what can be learned about the decision-making process of officers using deadly force.

Thus, a bottleneck exists in understanding officer-involved shootings. On the one hand, controlled laboratory tasks ask a specific question that may miss out on a broader understanding of how officers decide to shoot. On the other hand, archival data analyses suggest that situational factors removed by controlled laboratory tasks play a central role in the decision to shoot but cannot precisely describe how these factors impact officers' decision processes. Neither approach is especially well-suited to recommending training interventions to reduce fatal shootings or racial disparities therein.

Our approach was to recruit a large sample of officers to make decisions in a shooting simulator similar to those used in law enforcement training (*Fig 1*). We created the Immersive Shooting Simulator (ISS) to study officers' deadly force decisions in a repeatable, controlled situation that captures the richer policing scenario surrounding the decision to shoot. During the ISS, officers interacted with suspects according to protocol in life-sized videos shot from a first-person point of view. If the officers decided to use deadly force, they used a modified handgun to shoot. The gun fired with realistic sound and recoil and recorded response times. Scenarios were developed with input from the police and depicted suspects across various scenarios (e.g., traffic stops, arrest warrants). This approach allowed us to quantify the degree to which suspect race and policing scenario independently contributed to officers' decisions to shoot, and use cognitive modeling to describe the underlying cognitive process as officers made these decisions (described in more detail below).

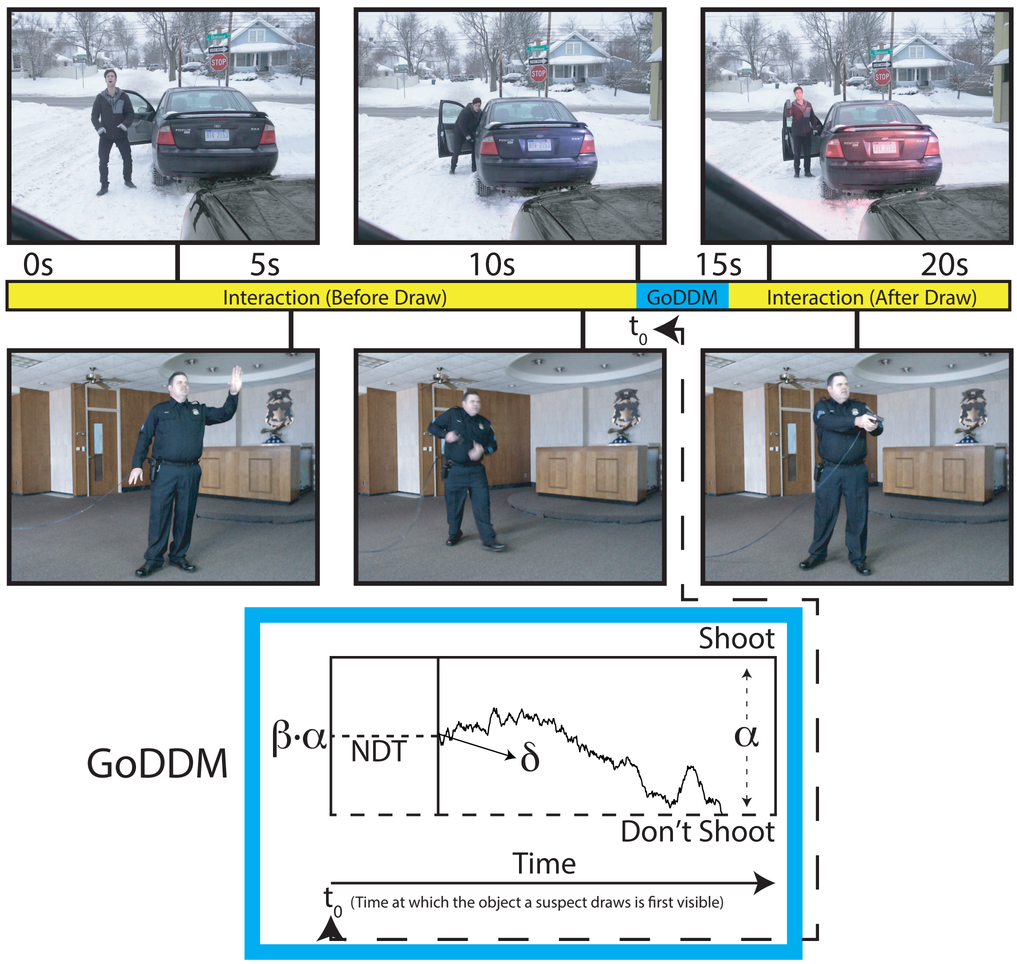

Fig 1. A trial from the ISS. In this scenario, the officer pulls over a suspect for speeding. The officer talks to the suspect until he reaches into his car, at which time the officer draws his weapon. At about 13 seconds, the critical object (a cell phone) is first visible (t_0). The shooting decision process was modeled from the point the object is drawn using the GoDDM (pictured at the bottom) until the decision was made. In this case, the suspect drew a cell phone and the officer did not shoot.

By moving to a shooting simulator with videos of real policing scenarios, many potential questions arise about how various factors of the scenarios might impact the decision. Here, we focus on one question: suspect behavior. We assigned four actors (two White men and two Black men) to each scenario. Within each scenario, half the actors were instructed to act antagonistically toward the officer, and half were instructed to act nonantagonistically or in a compliant manner. This decision reflects the concern in prior research, particularly in the criminal justice literature, about suspect behavior and escalation. It also reflects the broad concern about the most tragic type of fatal police shooting, the shooting of unarmed citizens who are not behaving in any aggressive manner (e.g., the fatal officer-involved shooting of Philandro Castile during a traffic stop 28; 29).

Finally, some past research has used similar shooting simulators (10; 30). Our work goes beyond this past research by (1) fully crossing suspect race and policing scenario, (2) modeling variation in policing scenario and suspect behavior, (3) modeling the underlying decision process, and (4) recruiting larger officer samples. These features allow us to more precisely estimate the effects of all possible sources of information on the underlying decision process.

Materials and methods

Participants

Sworn officers ($N = 659$) from the Milwaukee Police Department participated in the study in the Spring of 2017. The project was introduced each morning at officers' in-service as a study on expert decision-making, focusing on fast shoot/don't shoot decisions. We emphasized that officers would complete multiple scenarios but did not mention race as a factor we were studying. After this description, officers voluntarily signed up for individual time slots staggered throughout the day. No compensation for participation was given. We collected data from as many officers as possible over the 10-week period covering officers' spring in-service sessions. The department had a total sworn officer body of about 1,800 officers at the time, though in practice, not all 1,800 had the opportunity to participate as we were not present every day.

Officers completed the self-paced task individually. They saw up to 32 different scenarios. Most officers (75%) completed all trials, and almost all officers (89%) completed at least 24 trials. However, some officers completed fewer trials because they exceeded the 20-minute limit or experienced technical difficulties. We removed trials where officers responded before the object was first visible (60), responses that were more than three standard deviations above the mean response time for a scenario (127), or where the gun malfunctioned (97). In total, of the 19,600 total observations, we removed 284 trials (1.44%). Our analyses are based on the final sample of 19,316 trials. Analyses of the full data set led to the same conclusions. These regressions are available on the OSF site.

We visually determined that 592 (90%) of these officers were men and 484 (73%) were White, 103 (16%) were Black, 55 (8%) were Hispanic, 14 (2%) were Asian, and 3 (1%) were from other groups. Sample demographics were fairly representative of the department, although White officers were overrepresented (73% compared to 63% in the department). In total, 94% of the officers reported that they had an average of 11 years as a sworn officer ($SD = 7$, range $= 0 - 25$).

Procedures

Officers were told that the research aimed to understand how experts make fast decisions, particularly regarding object identification and the decision to shoot. Officers were told they would watch a series of policing scenarios and were instructed to interact with suspects as they would on the job. Officers were informed that if a suspect in the video pulled a gun, he would fire it at the officer, and they should fire the modified handgun at the suspect. Officers provided consent to have their data used for research purposes and were then asked to consent for their session to be filmed. Officers began each trial with the handgun holstered and read the dispatch information displayed on the screen at their own pace. When they indicated they had read the information, the trial began. After each trial, officers re-holstered their guns, and dispatch information for the next scenario was displayed. Officers were thanked and dismissed after completing all trials or the allotted twenty minutes.

Materials

Shooting simulator

Scenarios were displayed using a custom-built shooting simulator similar to commercial law enforcement training simulators. Videos were projected at near life-size. Officers began each trial roughly 15 feet from the screen and were encouraged to talk to suspects and move around as needed. Shooting responses in the simulator were made using a Glock handgun, modified with a Dvorak Air Recoil System. This system replaces the magazine and barrel of the handgun with a compressed CO_2 system, which cycles the gun as normal and provides recoil when the trigger is pulled. We further modified the system so that each trigger pull activated a microcontroller signaling to the computer that the trigger was pulled with near-millisecond accuracy. The signal prompted the computer to play the sound of a Glock handgun firing a live round through a set of speakers placed near the screen. All aspects of video presentation and response recording were controlled with PsychoPy (31). Detailed plans can be found at the OSF page.

Video scenarios

We collaborated with the Milwaukee Police Department to design and film a set of realistic scenarios commonly encountered by officers. We filmed eight scenarios (see full descriptions in the *Supplemental Online Materials*). Scenarios were filmed from the point of view of the officer and lasted around 20 seconds. All scenarios had a similar structure. After an initial interaction with a suspect there were two pivotal moments: One in which the suspect would perform an ambiguous action that raised the threat level for the officer (e.g., reaching into a glove box), and another in which the suspect would draw either a harmless object or a firearm. It was at this point officers had to decide to shoot. If the suspect in the video drew a firearm, he always shot at the officer. Officers were under time pressure—the suspect always fired the gun within one second after it was drawn. Although the specific draw time varied across scenarios, within each scenario, draw time was digitally manipulated to be equal within one video frame across suspect races.[1]

We employed ten Black male actors and ten White male actors as suspects. Each actor was filmed twice per scenario (across at most two different scenarios). In one video, the actor drew a handgun and fired at the officer; in the other, they revealed a harmless object such as a wallet or cellphone. Within a scenario, actors were matched in age, height, and clothing type, which was non-diagnostic of socioeconomic status.

We also manipulated the degree to which suspects escalated the interaction by acting antagonistic or non-antagonistic. Within each scenario, one version was produced

[1]The simulator was programmed so that if a participant completed the 32 trials, then there would be an equal number of trials (4) across the conditions formed by crossing the race by object by antagonistic manipulations. Due to a computer error, the first nine participants had five observations in the black, nongun, antagonistic condition.

with the actor acting in an antagonistic manner toward the police, and one version was produced with a different actor acting in a non-antagonistic manner toward the police. For example, in a pullover scenario, one version had an actor exiting his car swearing at the officer, and one had an actor exiting his car with his hands raised and pleading with the officer. This manipulation was fully crossed with race and armed status within each scenario.

Before each scenario, officers were given basic dispatch information about the reason for being at the scene. Dispatch information was randomly varied and blocked within the scenario. In all, we created 64 different videos from the eight scenarios. Each scenario had an antagonistic and non-antagonistic version, crossed by suspect race (White, Black) and armed status (unarmed, armed). Officers saw half of these videos, randomized such that officers could not predict whether a suspect was armed or not based on which versions of each scenario they had already seen.

Measures

On each trial, we recorded whether an officer fired the gun, the response time associated with the first shot, and the number of shots fired. For officers who consented to be filmed (95%), we also coded whether and when they grabbed their weapons. For the weapon grab data, we only coded the first eight trials of the task to avoid anticipation effects that might occur when an officer sees the same scenario again later in the task. We were able to code 90% of the 4,058 taped trials. The remaining 10% of the trials could not be coded because officers stepped out of frame, the video was too dark, or there were technical difficulties, leaving 3,656 trials.

Given that past research on the decision to shoot using controlled laboratory tasks has focused on the observed decision and the response time associated with the shooting, we present full analyses of these measures here in the main text. More detailed analyses involving the other measures are presented in the *Supplemental Online Materials*.

Analytic Approach

Behavioral Modeling

Decisions 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 conducted a model comparison between different models, identifying the model that best performed according to leave-one-out (LOO) cross-validation (see *Supplemental Online Material*). For the error rates, the best-performing model had random intercepts for scenarios and suspects and random slopes for the conditions nested within scenarios. For response times, the best-performing model had random intercepts for each unique scenario by suspect behavior combination (i.e., video). But, the model with random intercepts for scenarios and suspects and random slopes for the conditions nested within scenarios had a nearly indistinguishable fit (i.e., within the margin of error) and provided the same conclusions. Thus, for all behavioral modeling in the main paper, we report the model with random intercepts for scenarios and suspects and random slopes for the conditions nested within scenarios. In the main paper, we report the credible effects and the credible contrasts. The OSF site reports summaries of each of the estimated models. The summaries include estimates of the posterior distribution of the regression coefficients, estimates of the mean error rates, and response times for each manipulated condition. We also include estimates of the race effect for the object by suspect behavior condition.

These multilevel models were estimated with Markov Chain Monte Carlo (MCMC) methods implemented via the rstanarm package (32) in R. This program enables full Bayesian statistical inference using an estimation approach (33). We ran four chains using the MCMC sampler to draw from posterior distributions of parameters, with 9000 samples per chain (to ensure an effective sample size of >10000 for each coefficient), and a burn-in of 1000 samples. We investigated the convergence of posteriors through visual inspection and the Gelman-Rubin statistic (34). We used the default weakly informative priors in rstanarm (version 2.32.1). We report the posterior predicted mean of the parameter or statistic of interest and (in brackets) its 95% Highest Density Interval (HDI; 33). 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.

Cognitive Modeling

We used computational modeling to understand officers' underlying decision processes. A formal cognitive model uses mathematical language to specify how basic cognitive processes give rise to a phenomenon of interest (35; 36). This approach synthesizes these hypothetical processes in an observable and testable form. In this case, we used the Diffusion Decision Model (DDM; 37). The basic DDM models the decision process as an evidence accumulation process and decomposes this process into four psychological parameters (*Table 1*). According to the DDM, participants deciding whether to shoot or not begin with an initial bias toward one option or the other as indexed by the parameter β. This bias forms a start point from which participants begin accumulating evidence. They accumulate evidence by repeatedly sampling relevant information from the environment. The drift rate δ describes the average strength of the evidence in each sample, thus capturing the average rate at which evidence evolves toward the options. When the evidence reaches a threshold, the corresponding option is chosen. The parameter α describes the separation between the thresholds, indexing the amount of evidence required to make a decision. Finally, the model assumes there are contaminants to response times beyond the deliberation time captured by the evidence accumulation process. These contaminants are captured by the *Non-Decision Time* (NDT') parameter. The parameters of the model have been validated at the cognitive level (e.g., 38) and to some extent the neural level (for a review see 39). Moreover, the DDM has been established to accurately describe the decision to shoot in simplified laboratory shooting tasks (8; 11; 40).

One limitation of the DDM is that it is usually used to model two-choice tasks. Yet, in both the ISS and the field, officers register only one explicit response: to shoot. This different response mode does not necessarily imply a different process. Work in cognitive psychology has shown that when people make go/no-go decisions, they use the same evidence accumulation process as when making two-alternative forced-choice decisions. The no-go response (in this case, "Don't Shoot") is an implicit response (41; 42). That is, the "Don't Shoot" response is made at some point but is just not explicitly known. We incorporated this assumption into the model, treating the data as missing or censored and modeling this missing data (see the *Supplemental Online Material* for details). We refer to our revised model as the GoDDM. See *Figure 1* for an illustration of the GoDDM process and *Table 1* for a description of the model parameters.

We used hierarchical Bayesian methods to estimate the GoDDM (33; 43; 44). This approach simultaneously models individual and group level parameters, making it possible to estimate the model in this dataset where a larger number of officers complete a smaller number of trials. This approach also allowed us to model the

Table 1. Parameters of the Go Diffusion Decision Model and Their Substantive Interpretations.

Drift Diffusion Model Parameter	Description
Relative start point (β)	The location of the starting point for evidence accumulation relative to the thresholds, with $0 < \beta < 1$. The relative start point indexes an initial bias for either response, with values of β greater than .5 indicating a bias to choose "Shoot" and values lower than .5 indicating a bias to not shoot.
Threshold separation (α)	The separation between the thresholds, with $0 < \alpha$. With this parameterization, the choice threshold for "Shoot" is set at α, and the choice threshold for the "Don't Shoot" (unobserved) option set at 0. The threshold separation determines how much a person trades accuracy for speed (i.e., the speed–accuracy tradeoff), with larger values indicating more accurate but slower decisions.
Drift rate (δ)	The average strength in evidence at each unit of time, with $-\infty < \delta < \infty$. The sign of the drift rate indicates the average direction of the incoming evidence, with negative values indicating evidence in favor of "Don't Shoot" and positive values indicating evidence in favor of "Shoot." The magnitude of the drift rate characterizes the quality of the incoming information.
Relative non-decision time (NDT')	Proportion of response time (relative to the minimum observed response time) spent on processes unrelated to decision-making, with $0 < \mathrm{NDT}' < 1$. The non-decision time includes the time spent on encoding the stimulus, executing a response, and any other contaminant process.

implicit "Don't Shoot" response by imputing missing response time data.

Parameter recovery Model recovery analyses confirmed that the Bayesian framework and experimental design allowed for relatively accurate recovery of GoDDM parameters (see *Supplemental Online Material*). Due to the smaller sample size per condition, the recovery analyses suggest a small underestimation of the relative start point ($\sim .03$) and 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 imprecision, particularly in the threshold separation, appears to arise from using a random-walk approximation of the diffusion process to simulate the models; the other part of the imprecision is due to the small samples per condition at the individual level. We designed the recovery analysis to also ask if there was an effect of race in a subset of conditions (e.g., nonantagonistic), would we be able to recover the difference? We ran two models, one with a race effect in the relative start point and another model with a race effect in the drift rate. In both cases, we coded the effect to be about 0.75 difference in terms of standardized mean difference (in terms of between-participant standard deviation). In both cases, we could recover the difference with a high degree of accuracy. Overall, the accuracy of the parameter recovery — particularly between conditions — shows that the order and magnitude of the parameters were maintained. 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) (45; 46). Thus, even if a credible effect may not be found the values of the parameters carry information speaking to the effect.

Model comparison We parameterized the GoDDM to investigate how object, suspect race, and suspect behavior impacted each model parameter (*Table 1*). We started with a base model that allowed the start point and threshold to vary by race and the drift and non-decision time to vary by race and object. All the parameters modeled participant-level variability as a random intercept. This base model reflects the model that has the best-modeled choice and response times in the laboratory-based first-person shooter tasks (8; 47). Our behavioral analyses highlighted the importance of modeling the heterogeneity in behavior due to differences in the scenario and suspect. 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. The full comparison is reported in the Supplementary Material. Briefly, We compared several different GoDDM models, asking what parameters these situational factors primarily influenced. We made model comparisons using the Deviance Information Criterion (DIC; 48). This information criterion metric is useful for hierarchical models and penalizes complexity models. Focusing first on the effect of the scenario, the best model isolated the effect of the scenario on the relative start point. In terms of suspect effect, the best model isolated the effect of the suspect on the relative start point. We then coded a final model that had variability in the relative start point for both the scenario and suspects (as well as officers). This model had a worse fit (according to DIC) than the model with variability in the start point for suspects and not the scenarios. Nevertheless, we used the full model with variability in the scenario in the relative start point for both scenarios 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 the scenario contributes to this variability.

Results

We analyzed our data at the behavioral and cognitive levels. The behavioral level focuses on error rates and response times for armed suspects. The cognitive model uses the GoDDM model to estimate the process-level parameters that generate the choices (i.e., errors) and response times.

Behavioral Analyses

We start by analyzing the behavioral data of errors and response times in the decision to shoot. We analyze error rates and responses with Bayesian hierarchical models. We model the variability between participants by including random intercepts for officers. Model comparisons (see Analysis; Supplementary Material) reveal that the most useful models for accounting for scenarios and suspects were models with random intercepts for scenarios and suspects and random slopes for the conditions nested within scenarios.

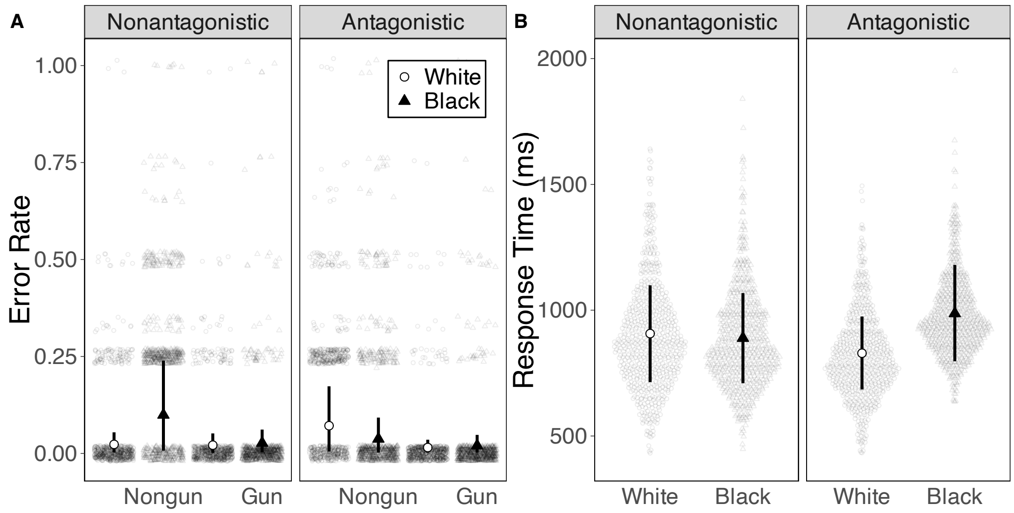

Fig 2. (A) Effect of object and race on errors for nonantagonistic and antagonistic suspects at the group level (solid black outline) and individual level (light grey). (B) Effect of object and race on response times for nonantagonistic and antagonistic suspects at the group level (solid black outline) and individual level (light grey). Non-greyed markers indicate mean posterior estimate and bars indicate 95% HDI. All group-level estimates were calculated from the group-level posterior predicted distributions from the hierarchical models.

Error rates

Figure 2A shows the error rate by suspect race, presence of weapon, and suspect behavior. There was a credible three-way interaction between suspect race, weapon presence, and suspect behavior $b = 2.21$, [0.37, 4.09]. This interaction was driven by the effect of race on unarmed, nonantagonistic suspects. Unarmed nonantagonistic Black suspects were more likely to be shot than unarmed nonantagonistic White suspects ($M = 1.52$, [0.17, 2.88]), but there was no credible difference between Black and White nonantagonistic armed suspects or when the suspects were antagonistic.

At first glance, this three-way interaction between race, object, and suspect behavior appears consistent with past work where the decision to shoot is based on threat perception and race provides information about threat under ambiguous circumstances (7). When the suspect acts nonantagonistically, negative stereotypes of race may drive this threat perception; in the antagonistic condition, the behavior determines threat perception, minimizing the effect of race. However, there was not a credible increase in error rates in the antagonistic ($M = 2.7\%$, [0.7,5.2]) vs nonantagonistic condition ($M = 3.1\%$, [0.9,6.0]; $b = 0.137$ [-0.771, 1.017]). Moreover, exploring the data further suggests that part of this interaction may be due to particular behaviors of some actors in specific scenarios. We explore this speculation further, but first, we report the corresponding analyses with response times.

Response times

Figure 2C displays officers' response times to shoot armed suspects by suspect race for nonantagonistic and antagonistic behavior. There was no credible effects of the manipulations on the response times.

Additional behavioral measures

We measured several other behaviors during the immersive shooting simulator. These behaviors were the number of shots fired, whether the officer reached or grabbed for the gun, and the time at which they reached for the gun. We report these results in the Supplementary Material. We only found a credible effect of race in terms of the number of shots fired. A Poisson regression revealed 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]). There was no credible effect of race for nonantagonistic armed and antagonistic suspects.

Scenario and suspect effects

We sought to understand the interactions between race, object, and suspect behavior across the scenarios and suspects. Figure 3 plots the error rates and response times at the level of the scenarios for each condition. Figure 4 does the same for each suspect. As the figures show, there was a fair amount of heterogeneity at the scenario level and the suspect. Regarding error rates, Figure 3 shows the effect for unarmed, non-antagonistic Black suspects isolated primarily to the alley and pullover scenarios. In these videos, the actors draw their wallets in a manner that mimics the drawing of a gun. Figure 4 shows the heterogeneity at the suspect level. Unfortunately, the experimental materials make it difficult to separate the suspect from the scenario as a given suspect was in at most two scenarios.[2]

As described earlier, there is evidence that real-world decisions to shoot are influenced by situational factors. The ISS makes it possible to quantify how much the decision to use deadly force is associated with the situation or suspect behavior. We tested this by calculating intra-class correlations (ICCs) to assess how much variation in behavior was associated with officers, policing scenarios, and suspects in general. Variability in the decision to shoot was primarily associated with scenarios ($M = .19$ [.05, .38]) and suspects ($M = .12$ [.05, .22]). Less variability was associated with officers ($M = .07$ [.04, .09]). In contrast, for shooting response times, variability was primarily associated with officers ($M = .19$ [.15, .24]) and scenarios ($M = .14$ [.03, .30]), rather than suspects ($M = .11$ [.05, .19]); some officers were faster to shoot than others.

Summary

Overall, the behavioral results reveal that many features of the decision situation, including policing scenario, suspect race, and suspect behavior, impact the behavioral decision to shoot. We next used cognitive modeling to understand how these factors impacted the decision process as measured by the parameters of the GoDDM (see *Table 1*).

Cognitive Modeling: GoDDM Analyses

To examine the cognitive processes underlying officers' deadly force decisions, we adapted the Bayesian Diffusion Decision Model (DDM) framework used for the first-person shooter task (8; 11) for the immersive shooting simulator. See *Figure 1* for an illustration of the GoDDM process and *Table 1* for a description of the model

[2]For the alley and night pullover scenarios, the Black actors only appeared in those scenarios (suspects 3 and 4 in the alley scenario, suspects 1 and 2 in the pullover scenario). In the alley condition the White suspects only appeared in that scenario.

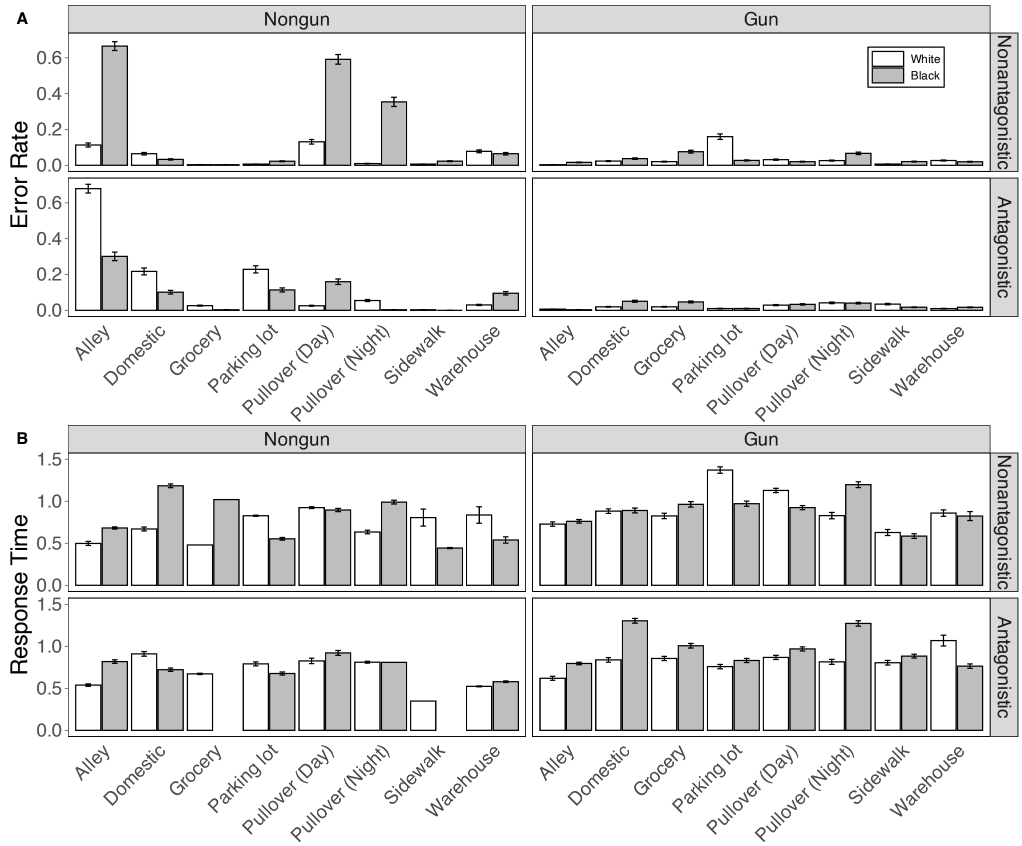

Fig 3. (A) Error rates by scenario split by whether the suspect was unarmed or armed and acting nonantagonistically or antagonistically. (B) Mean response times by scenario split by whether the suspect was unarmed or armed and acting nonantagonistically or antagonistically. Error bars are 95% confidence intervals.

parameters. The decision to shoot is a go/no-go decision where participants must respond to one option (shoot) but withhold their response to another (don't shoot) (49). Extensive modeling work has shown that people in a go/no-go procedure use the same evidence accumulation process as when they complete a two-alternative forced-choice task (41; 42). The difference is that the no-go response (i.e., "Don't Shoot") is implicit (41). 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 (33). Model recovery analyses confirmed that the model parameters can be accurately recovered for this adaptation of the DDM, the GoDDM, with the experimental design used in this study (see *Supplemental Online Material*).

The GoDDM was parameterized to examine how suspect race, suspect behavior, and their interaction impacted each parameter. We modeled participant-level variability by treating participants as a random intercept for each process parameter. Our behavioral analyses highlighted the importance of modeling the heterogeneity in behavior due to differences in the scenario and suspect. 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

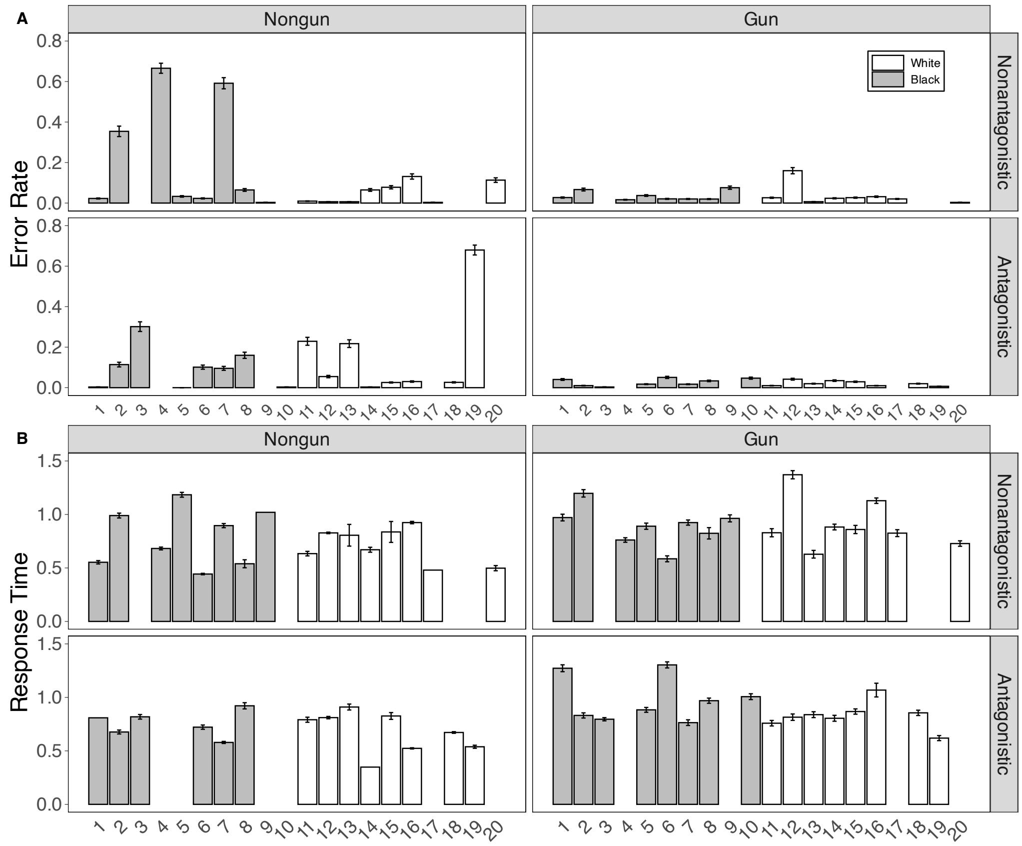

Fig 4. (A) Error rates by suspect split by whether the suspect was unarmed or armed and acting nonantagonistically or antagonistically. (B) Mean response times by split by whether the suspect was unarmed or armed and acting nonantagonistically or antagonistically. Error bars are 95% confidence intervals.

comparison to ask in what process parameter the effect of scenario and suspect should be isolated. The model comparison revealed that the effect of scenario and suspect was in terms of the relative starting point. We use this full model to examine the manipulations of race and suspect behavior on the process parameters.

Process parameters

Figure 5 plots the group-level estimates for each of the parameters of the GoDDM by the race of the suspect and whether the suspect behaved antagonistically. One key aspect that this cognitive-level analysis highlights is that the effects of race and suspect behavior do not have selective influence on a single parameter in the GoDDM. Instead these factors give rise to a constellation of effects on the decision process of officers.

Relative start point In terms of the relative start point β, officers showed no initial bias to shoot Black suspects, M = -0.005 [-0.037, 0.028] (*Figure 5A*). In contrast, there was a credible effect of the suspect's antagonism on the initial bias, with officers starting closer to "Don't Shoot" for antagonistic versus nonantagonistic suspects, M = -0.056 [-0.073, -0.037]. It could be that this drop in the relative start-point reflects officers trying to be strategically hesitant to shoot with suspects who are being antagonistic but, as of yet, have not pulled an object. With respect to the race effect found in the behavioral analyses, these results suggest that increased errors to shoot

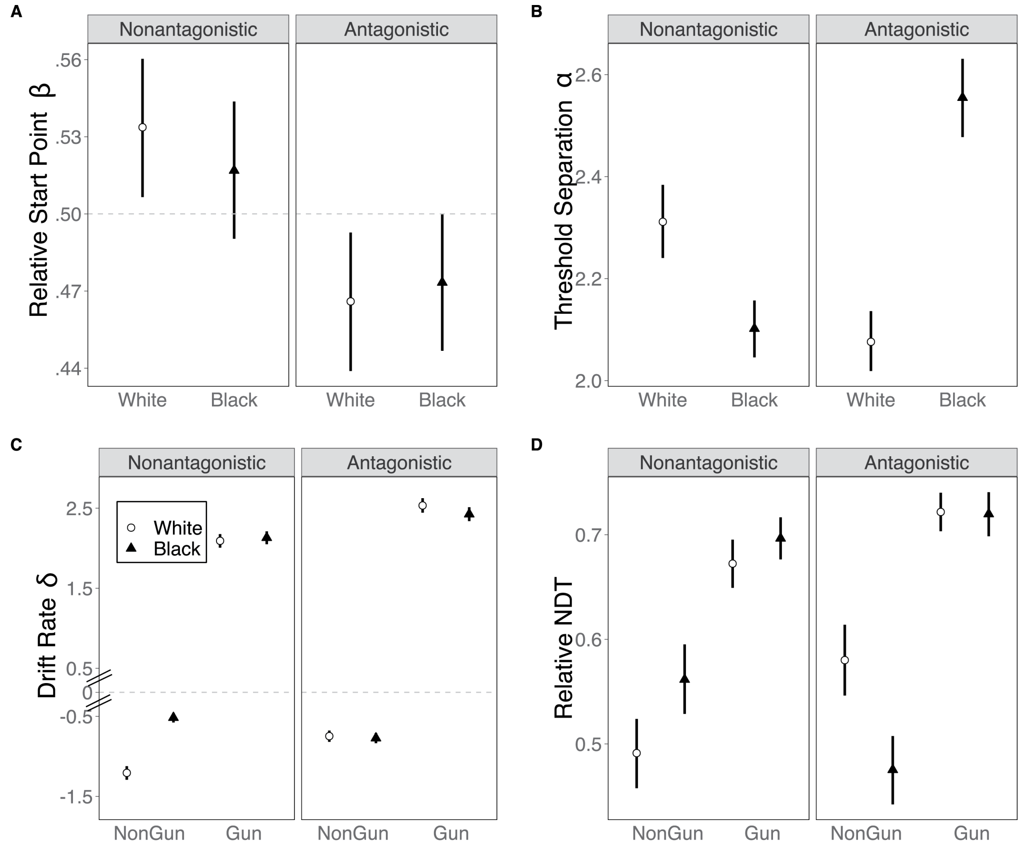

Fig 5. Diffusion model parameters. (A) Group level estimates of relative start point by race for nonantagonistic and antagonistic suspects. (B) Group level estimates of threshold separation by race for nonantagonistic and antagonistic suspects. (C) Group level estimates of drift rate by object type for White and Black suspects and nonantagonistic and antagonistic suspects. (D) Group level estimates of relative non-decision time by object type for White and Black suspects and nonantagonistic and antagonistic suspects. Markers indicate mean posterior estimate and bars indicate 95% HDI.

Black unarmed, nonantagonistic suspects did *not* result from officers having an initial bias to shoot Black suspects before the moment the object was visible. This lack of an effect also is inconsistent with the explanation that race impacted threat assessment throughout the interaction in the nonantagonistic condition by pushing officers closer to the "Shoot" decision prior to the object being presented.

Threshold separation *Figure 5B* summarizes the posterior estimates of the group-level threshold separation parameter. This parameter measures the amount of evidence officers sought to collect and thus can index their overall caution. It shows a credible interaction between race and suspect behavior, M = -0.344 [-0.408, -0.278]. Specifically, compared to nonantagonistic White suspects, officers had a credibly lower threshold separation for nonantatonistic Black suspects, M = -0.210 [-0.298, -0.121]. Conversely, for antagonistic suspects, officers had a credibly higher threshold separation for antagonistic Black suspects, M = 0.479 [0.385, 0.574]. Taken together, this pattern implies that how cautious officers were with suspects during the task depended on their race and how the suspects behaved, with officers showing more caution with Black suspects acting antagonistically and less caution with Black suspects behaving non-antagonistically. At the behavioral level, while not a credible difference, this explains the slower responses officers tended to have for armed Black antagonistic suspects ($M = 989$ ms [762,1,222]) compared to White antagonistic suspects ($M = 825$ ms [661,996]).

Drift rate *Figure 5C* summarizes the posterior estimates of the group-level drift-rate parameter. This parameter captures the rate and direction of evidence accumulation indexing the strength of the evidence towards shooting or not that officers extracted from the scene.

The effects of race and suspect behavior on the drift rate correspond closely to the observed error rates. In particular, there was a credible three-way interaction between race, object, and suspect behavior, $M =$ -0.216 [-0.276, -0.156]. This three-way interaction was due to officers showing a stronger drift rate towards *not shooting* unarmed nonantagonistic White suspects compared to the corresponding Black suspects, $M =$ -0.690 [-0.794, -0.587]. Further comparisons revealed that, relative to the antagonistic counterparts, there were two distinct effects for White and Black suspects. For White suspects, officers showed stronger drift rates toward not shooting unarmed non-antagonistic suspects vs. antagonistic suspects ($M =$ -0.447 [-0.544, -0.352]). Conversely, for Black suspects, officers showed weaker drift rates toward not shooting unarmed non-antagonistic suspects vs. antagonistic suspects ($M = 0.243$ [0.167, 0.318]).

Notably, this difference between nonantogonistic and antagonistic unarmed suspects is almost twice as large for Whites vs. Blacks ($M = 1.90$ [1.13, 2.77]). This is consistent with race being used as a cue in deciding to shoot, perhaps as a way to disambiguate an ambiguous situation. More broadly, this effect of race in the drift rates combined with the decreased threshold separation for Black nonantagonistic suspects can explain the increased error rate for these suspects. There were no other credible differences between the drift rates.

Non-decision time *Figure 5D* summarizes the posterior estimates of the group-level non-decision time parameters. This parameter captures the amount of time relative to the smallest observed response time that was due to other processes besides deliberation such as visual search or response selection. The estimates show that response times had greater levels of contaminants for Black (vs. White) unarmed suspects behaving nonantagonistically, $M = 0.07$ [0.02, 0.12]. But, the opposite was true of unarmed antagonistic suspects, $M =$ -0.10 [-0.16, -0.06].

Cognitive Modeling: Scenario and Suspect Effects

As we discussed, the model accounted for scenario and suspect effects as impacting the relative start point. Figure 6 plots the relative start points across the scenarios. These estimates reveal that officers showed a credible bias to shoot in scenarios where there was a greater risk of threat in the period leading up to the decision to shoot. This included scenarios where suspects had a warrant for their arrest (alley and sidewalk scenarios) and the only scenario where a suspect was carrying a non-gun object that could be used as a weapon (a crowbar used in the warehouse scenario). The GoDDM data map onto the behavioral data; officers were faster to shoot armed suspects in the warehouse and sidewalk scenarios than in other scenarios and more errors in the alley (Figure 2).

The relative start points for each suspect are plotted in Figure 7 when the actor was acting antagonistically or not. This figure emphasizes the amount of variability at the process level that is driven by the different suspects (independent of race and whether

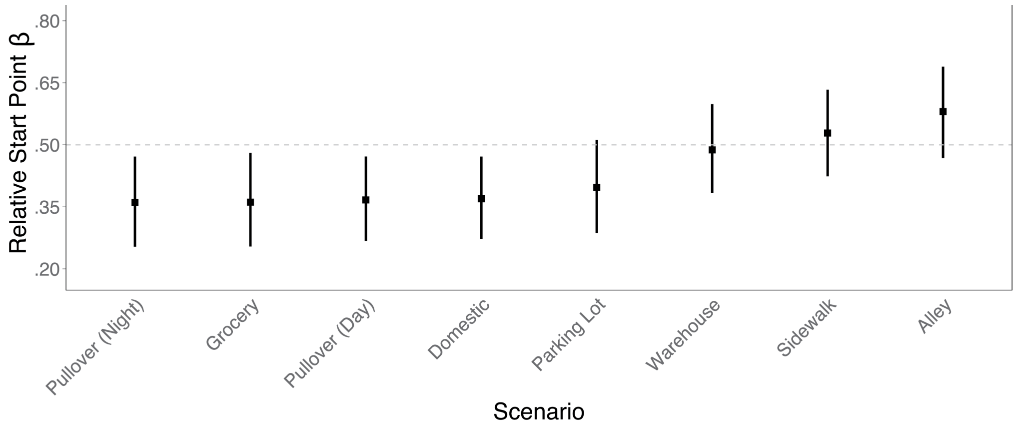

Fig 6. Relative start points for each of the scenarios. Dots indicate mean posterior estimate and bars indicate 95% HDI.

they behaved antagonistically or not). In fact, one can see just how much more variability there is between suspects in terms of the relative start points they engendered as compared to the other manipulations.

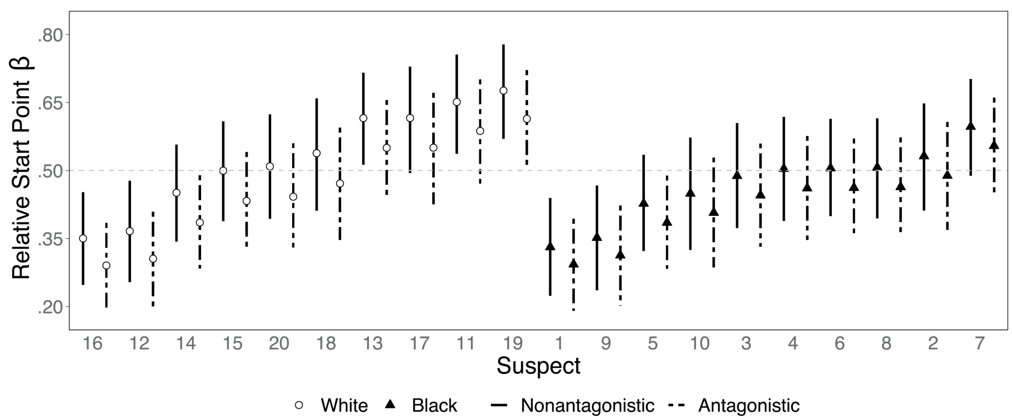

Fig 7. Relative start points for each of the suspects. Markers indicate mean posterior estimate and bars indicate 95% HDI.

Summary

Taken together, the GoDDM helps isolate at the cognitive level how the numerous factors influencing the behavioral decision to shoot enter the decision process. Analyses suggest that when race enters the decision process, it is not via changes to the initial proclivity to shoot but as information that is accumulated as officers decide to shoot. But, this use of race as information was limited to a specific condition: when a suspect was acting in a nonantagonistic manner. At the same time, officers appear to respond to the race of the suspect in the ISS by adjusting their threshold widening their threshold separation in a very charged situation when the suspect was acting antagonistically. This combination of effects explains the increased error rate for unarmed Black suspects acting nonantagonistically and slower response times for Black suspects acting antagonistically. At the same time, the GoDDM also helps model how situational factors like the scenario and the suspect's general characteristics enter the decision process. In this case, these factors change the relative start point or the officer's initial proclivity to shoot.

Discussion

We used an immersive shooting simulator and cognitive modeling to understand officers' deadly force decisions. This unique combination allowed us to investigate how policing scenario, suspect behavior, and race impact the decision process.

The behavioral analyses suggest the following primary conclusions. First, a three-way interaction between race, object, and antagonistic behavior was observed such that officers were more likely to incorrectly shoot unarmed Black suspects acting nonantagonistically relative to the corresponding White suspect. We also observed a similar effect regarding the number of shots fired. There were no effects of race or suspect behavior on the response times. As we discuss next, our computational cognitive model—the GoDDM—isolates the cognitive mechanism driving this result to the information officers extracted from the scene during these trials. Importantly, we also found that across all the trials, there was substantial variability in shooting decisions related to scenarios and suspects, with the race effect varying across specific scenarios and actors.

The process-level analyses provide a deeper understanding of the decision to shoot in a way that is not obvious from the behavioral data alone. Although increased errors in shooting unarmed suspects might have reasonably been predicted due to officers being more likely to shoot before encountering the object, the DDM analysis isolates this effect to the type of information accumulated (i.e., the drift rate) and how much information was accumulated (i.e., the threshold separation). Specifically, for nonantagonistic suspects, drift rates (i.e., the rate of evidence accumulation) pointed more strongly toward not shooting unarmed White suspects and more strongly toward shooting unarmed Black suspects. Combined with the lower threshold for accumulating evidence for unarmed, nonaggressing Black suspects, this resulted in increased errors in these conditions.

At the same time, behavior in the ISS also revealed the considerable role that policing scenario and suspect behavior had on officers' decisions to shoot. This ability to study the decision to shoot during the more complex and realistic policing scenario connects the existing psychology and criminal justice literature in a way that fills in the weaknesses inherent in each research tradition. The ISS also expands beyond these findings as it helps isolate at the process level how scenario and acting behavior enter the decision. In particular, information about the policing scenario and suspect behavior enter the decision to shoot via officers' initial starting point of evidence (via their relative start point) to be more likely to start the decision process closer to shoot or not shooting.

Effects of Race on the Decision to Shoot

A distinct advantage of the ISS and the GoDDM is that they permit direct tests of different hypotheses about how race (as well as other factors) enter the decision process. Several hypotheses can be ruled out based on this framework. For instance, a common hypothesis is that officers might have an initial bias to shoot a Black citizen; that is, they might have a higher initial start point (β) to shoot Black suspects. The results from the GoDDM do not support this hypothesis. Race did not impact the starting point from which officers began to accumulate evidence, and if anything, officers shied closer to an initial proclivity to not shoot for Black suspects.

Instead, the behavioral data in which officers were more likely to incorrectly shoot unarmed, nonantagonistic Black suspects compared to White suspects is best explained by the effect of race on the evidence accumulation process δ. This was realized in a stronger drift rate toward not shooting for unarmed, nonantagonistic White suspects and a stronger drift rate toward shooing for unarmed, nonantagonistic Black suspects (though still, on average, accumulating evidence toward not shooting). One possible reason for this decrease could be perhaps because officers focused on other aspects besides the object suspects held when those suspects were Black and nonaggressive. Another possible reason — not necessarily inconsistent with this first reason — is that stereotypes about Black males as violent may have made it more difficult to identify the nongun object when held by a Black suspect. This is consistent with the literature on stereotyping effects showing that stereotypes impact decisions only under conditions of ambiguity and not when there is clear diagnostic information, as in the case when the suspect is behaving antagonistically (50; 51; 52). A final possibility is that the drift rate effect (and the corresponding behavioral effect) was driven by a small number of actors behaving in specific ways. For instance, the videos in our study with a higher error rate had actors that appeared to draw harmless objects in a manner similar to pulling a gun. Further research is needed to understand this effect on drift rate more fully.

One might also speculate that the experimental context caused officers to feel observed and adjust their decision processes, particularly for Black suspects. This is certainly a possibility. According to the GoDDM, this would have been realized in one of two ways. One way is for officers to have become hypervigilant for Black suspects, increasing the sensitivity of their evidence accumulation for Black suspects (i.e., the difference in drift rates (δ) between armed and unarmed suspects). This was not observed. Another way officers might have responded to being watched is to increase their response caution and collect more evidence for Black suspects. Yet this would predict greater threshold separation (α) for Black suspects overall, whereas the race effect appears to have differed depending on the antagonism of the suspect. Moreover, threshold separation was *largest* when Black suspects were antagonistic. If officers were worried about appearing biased by shooting Black suspects, this concern likely would have manifested in the *nonantagonistic* cases; in the antagonistic cases, we suggest that officers would have been less likely to feel concerned over incorrectly shooting a Black target given that the aggression of the target would have provided a rationale for their decision.

Limitations

The ISS is, of course, still a simulation. Officers know they will not die or face formal consequences for incorrect decisions. Response options are also limited in the ISS, and suspects cannot respond to officers. Officers' actions (yelling at suspects) and physiological responses (sweating) suggest they were invested in the experience and took it seriously. Our results align with findings that situational and suspect-based measures are strong predictors of police behavior (25; 53; 54; 55; 56; 57). Yet it is important to acknowledge that this simulated method could measure a cognitive process distorted or not represented in actual, more variable shooting situations.

We urge caution in generalizing these findings. We sampled a single department, and nationwide variability exists in department policies and officer-involved shootings (27; 58). Our sample also may be different from the general population of officers. Although officers did not know the specific topic of study, self-selection is nevertheless possible (59). Moreover, the data were collected in 2017. The worldwide Black Lives Matter protests on police brutality (60; 61) that followed the killing of George Floyd (62) has led to some reform (63), the consequences of which are not well established yet. Similarly, we urge caution in generalizing beyond the specific scenarios used in our version of the ISS. Just as with actual-officer-involved shootings, each scenario is unique, and factors present in one situation may have different meanings in another. The fact that the ISS makes it possible to observe how many different officers respond to the same scenario makes it possible to extract general behavioral properties from a given scenario. Still, a richer theory about policing scenarios is needed to generalize to other scenarios.

In addition, we focus specifically on the decision to shoot; our findings may not generalize to other uses of force. A modest amount of the many policing situations that occur involve armed citizens, and even within this subset, not all armed citizens are deadly threats. Moreover, officers can respond to threats with various options, not just deadly force. Here we focus on the specific case where officers have to identify objects in the context of deadly force decisions quickly. Our conclusions should be constrained to these situations. Finally, we used computational modeling to investigate cognitive processes. This is one powerful method to get at the process level, but there are other approaches, including neural (e.g., 64), eye tracking (e.g., 40), and think-aloud or protocol analyses (65) that could also be used to study these critical decisions.

Conclusion

We used an immersive shooting simulator to understand officers' deadly force decisions and to investigate how race, variation in policing scenario, and suspect behavior impact the decision process. The ISS can advance our understanding of fatal police shootings by combining the control and precision of standard laboratory tasks with the policing variables known to be important from actual officer-involved shootings. Of importance, a suspect's behavior and policing scenario had large effects on the decision to shoot, and these factors exerted their influence at the process level by affecting officers initial proclivity to shoot. **??**.

Supporting information

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

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

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

S4 Fig. Observed and posterior predicted proportion of trials an officer shot in each condition. The observed average proportion are the $\times$, 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.

S5 Fig. 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.

S1 Table. Suspect race and scenario appearance..

S2 Table. Comparison of models of different models of scenario and actor variability for the error rates.

S3 Table. Comparison of models of different models of scenario and actor variability for the response times of armed suspects.

S4 Table. Intra-class correlations for outcome variables.

S5 Table. Summary of the parameter recovery analysis for two versions of the GoDDM.

S6 Table. DIC for models with different effects of scenario and suspect on the process parameters.

Acknowledgments

This work was supported by National Science Foundation Grant No. 1230281 to J.C. and Grant No. 1756092 to J.C. and T.J.P. David Johnson received funding as a postdoctoral researcher at the Lab for Applied Social Science Research at the University of Maryland. Funding was provided by the following Michigan State University offices: VP for Research and Graduate Studies, College of Social Science, Department of Psychology, University Outreach and Engagement, and Office for Inclusion and Intercultural Initiatives. We thank officers from the Milwaukee Police Department for their participation. This work would not have been possible without support from Chief Edward Flynn, Captain Nicole Davila, Lieutenant Richard Stein, Sergeant Paul Riehle, Sergeant Jeffrey Sunn, Larry Lumb from Saturn Manufacturing, David McFarlane, and Ahptic Film and Digital. We also thank the team of research assistants who helped collect data. Anonymous OSF link
https://osf.io/bgykz/?view$_{o}nly = 4ec1fa48d1e34dfbb8545cffcd95da6b$

Author Contributions

According to categories listed by Brand et al. (2015):

Conceptualization: J.C., D.J.J, & T.J.P.;
Methodology: J.C., D.J.J., & T.J.P.;
Software: J.C.;
Formal Analysis: T.J.P. & D.J.J.;
Investigation: J.C., D.J.J, & G.G.;
Data Curation: T.J.P . & D.J.J.
Original Draft: J.C., D.J.J. & T.J.P;
Reviewing & Editing: all authors;
Supervision: J.C.; T.J.P.
Project Administration: J.C., D.J.J., G.G, T.J.P.

References

1. Cesario J. What can experimental studies of bias tell us about real-world group disparities? Behavioral and Brain Sciences. 2022;45:1–71.

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

3. Payne BK, Correll J. Race, weapons, and the perception of threat. vol. 62 of Advances in Experimental Social Psychology. Academic Press; 2020. p. 1 – 50.

4. Correll J, Park B, Judd CM, Wittenbrink B. The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. Journal of Personality and Social Psychology. 2002;83(6):1314–1329. doi:10.1037/0022-3514.83.6.1314.

5. Correll J, Park B, Judd CM, Wittenbrink B, Sadler MS, Keesee T. Across the thin blue line: Police officers and racial bias in the decision to shoot. Journal of Personality and Social Psychology. 2007;92(6):1006–1023. doi:10.1037/0022-3514.92.6.1006.

6. Correll J, Park B, Judd CM, Wittenbrink B. The influence of stereotypes on decisions to shoot. European Journal of Social Psychology. 2007;37(6):1102–1117. doi:10.1002/ejsp.450.

7. Correll J, Wittenbrink B, Park B, Judd CM, Goyle A. Dangerous enough: Moderating racial bias with contextual threat cues. Journal of Experimental Social Psychology. 2011;47(1):184–189. doi:10.1016/j.jesp.2010.08.017.

8. Pleskac TJ, Cesario J, Johnson DJ. How race affects evidence accumulation during the decision to shoot. Psychonomic Bulletin & Review. 2018;25:1301–1330. doi:10.3758/s13423-017-1369-6.

9. Cesario J, Carrillo A. Racial bias in police officer deadly force decisions: What has social cognition learned? In: Carlston D, Johnson K, Hugenberg K, editors. *The Oxford Handbook of Social Cognition*. 2nd ed. Oxford University Press; 2024. p. xxx–xxx.

10. Cox WT, Devine PG, Plant EA, Schwartz LL. Toward a comprehensive understanding of officers' shooting decisions: No simple answers to this complex problem. Basic and Applied Social Psychology. 2014;36(4):356–364. doi:10.1080/01973533.2014.923312.

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

12. Plant EA, Peruche BM. The consequences of race for police officers' responses to criminal suspects. Psychological Science. 2005;16(3):180–183. doi:10.1111/j.0956-7976.2005.00800.x.

13. Sadler MS, Correll J, Park B, Judd CM. The World Is Not Black and White: Racial Bias in the Decision to Shoot in a Multiethnic Context. Journal of Social Issues. 2012;68(2):286–313. doi:10.1111/j.1540-4560.2012.01749.x.

14. Sim JJ, Correll J, Sadler MS. Understanding Police and Expert Performance: When Training Attenuates (vs. Exacerbates) Stereotypic Bias in the Decision to Shoot. Personality and Social Psychology Bulletin. 2013;39(3):291–304. doi:10.1177/0146167212473157.

15. Cesario J, Johnson DJ, Terrill W. Is there evidence of racial disparity in police use of deadly force? Analyses of officer-involved fatal shootings in 2015-2016. Social Psychological and Personality Science. 2019;10(5). doi:10.1177/1948550618775108.

16. Cesario J. On Selective Emphasis, Broad Agreement, and Future Directions: Reply to Ross, Winterhalder, McElreath. PsyArXiv. 2020;doi:10.31234/osf.io/2p5eg.

17. Fryer Jr RG. An empirical analysis of racial differences in police use of force. Journal of Political Economy. 2019;127(3):1210–1261.

18. Goff PA, Lloyd T, Gelle A, Raphael S, Glaser J. The science of justice: Race, arrests, and police use of force. University of California, Los Angeles: Center for Policing Equity; 2016.

19. Klinger D, Rosenfeld R, Isom D, Deckard M. Race, crime, and the micro-ecology of deadly force. Criminology & Public Policy. 2016;15(1):193–222. doi:10.1111/1745-9133.12174.

20. Mentch L. On Racial Disparities in Recent Fatal Police Shootings. Statistics and Public Policy. 2020;7(1):9–18.

21. Nix J, Campbell BA, Byers EH, Alpert GP. A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. Criminology & Public Policy. 2017;16(1):309–340. doi:10.1111/1745-9133.12269.

22. Ross CT, Winterhalder B, McElreath R. Racial Disparities in Police Use of Deadly Force Against Unarmed Individuals Persist After Appropriately Benchmarking Shooting Data on Violent Crime Rates. Social Psychological and Personality Science. in press;.

23. Scott K, Ma DS, Sadler MS, Correll J. A social scientific approach toward understanding racial disparities in police shooting: Data from the Department of Justice (1980–2000). Journal of Social Issues. 2017;73(4):701–722. doi:10.1111/josi.12243.

24. Tregle B, Nix J, Alpert GP. Disparity does not mean bias: making sense of observed racial disparities in fatal officer-involved shootings with multiple benchmarks. Journal of Crime and Justice. 2019;42(1):18–31.

25. Wheeler AP, Phillips SW, Worrall JL, Bishopp SA. What factors influence an officer's decision to shoot? The promise and limitations of using public data. Justice Research and Policy. 2017;18(1):48–76.

26. Worrall JL, Bishopp SA, Zinser SC, Wheeler AP, Phillips SW. Exploring bias in police shooting decisions with real shoot/don't shoot cases. Crime & Delinquency. 2018;64(9):1171–1192. doi:10.1177/0011128718756038.

27. Ross CT. A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011–2014. PLOS ONE. 2015;10(11):e0141854. doi:10.1371/journal.pone.0141854.

28. Pheifer P, Peck C. Aftermath of fatal Falcon Heights officer-involved shooting captured on video. Star Tribune. 2016;.

29. Domonoske C, Chappell B. Minnesota Gov. Calls Traffic Stop Shooting 'Absolutely Appalling At All Levels. NPR. 2016;.

30. James L, Klinger D, Vila B. Racial and ethnic bias in decisions to shoot seen through a stronger lens: Experimental results from high-fidelity laboratory simulations. Journal of Experimental Criminology. 2014;10(3):323–340. doi:10.1007/s11292-014-9204-9.

31. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. Behavior research methods. 2019; p. 1–9.

32. Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression modeling via Stan.; 2018. Available from: `http://mc-stan.org/`.

33. Kruschke J. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press; 2014.

34. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992;7(4):457–472.

35. Busemeyer JR, Diederich A. Cognitive Modeling. Sage; 2010.

36. Farrell S, Lewandowsky S. Computational Modeling of Cognition and Behavior. Cambridge University Press; 2018.

37. Ratcliff R, Smith PL, Brown SD, McKoon G. Diffusion decision model: Current issues and history. Trends in Cognitive Sciences. 2016;20(4):260–281. doi:10.1016/j.tics.2016.01.007.

38. Voss A, Rothermund K, Voss J. Interpreting the parameters of the diffusion model: An empirical validation. Memory & Cognition. 2004;32(7):1206–1220. doi:10.3758/BF03196893.

39. Gold JI, Shadlen MN. The neural basis of decision making. Annual Review of Neuroscience. 2007;30:535–574. doi:10.1146/annurev.neuro.29.051605.113038.

40. Correll J, Wittenbrink B, Crawford MT, Sadler MS. Stereotypic vision: How stereotypes disambiguate visual stimuli. Journal of Personality and Social Psychology. 2015;108(2):219–233. doi:10.1037/pspa0000015.

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

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

43. Lee MD, Wagenmakers EJ. Bayesian cognitive modeling: A practical course. Cambridge university press; 2014.

44. Vandekerckhove J, Tuerlinckx F, Lee MD. Hierarchical diffusion models for two-choice response times. Psychological methods. 2011;16(1):44–62. doi:10.1037/a0021765.

45. Gelman A, Tuerlinckx F. Type S error rates for classical and Bayesian single and multiple comparison procedures. Computational Statistics. 2000;15(3):373–390.

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

47. Johnson DJ, Tress T, Burkel N, Taylor C, Cesario J. Officer characteristics and racial disparities in fatal officer-involved shootings. Proceedings of the National Academy of Sciences. 2019;116(32):15877–15882.

48. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64(4):583–639. doi:10.1111/1467-9868.00353.

49. Donders FC. On the speed of mental processes. Acta psychologica. 1868/1969;30:412–431.

50. Jussim L. Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy. OUP USA; 2012.

51. Jussim L, Crawford JT, Anglin SM, Chambers JR, Stevens ST, Cohen F. Stereotype accuracy: One of the largest and most replicable effects in all of social psychology. Handbook of prejudice, stereotyping, and discrimination. 2016;2:31–63.

52. Kunda Z, Thagard P. Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. Psychological review. 1996;103(2):284.

53. Bolger PC. Just following orders: A meta-analysis of the correlates of American police officer use of force decisions. American Journal of Criminal Justice. 2015;40(3):466–492.

54. Council NR, et al. Fairness and effectiveness in policing: The evidence. National Academies Press; 2004.

55. Terrill W, Mastrofski SD. Situational and officer-based determinants of police coercion. Justice quarterly. 2002;19(2):215–248.

56. Terrill W, Reisig MD. Neighborhood context and police use of force. Journal of Research in Crime and Delinquency. 2003;40(3):291–321. doi:10.1177/0022427803253800.

57. White MD. Identifying situational predictors of police shootings using multivariate analysis. Policing: An International Journal of Police Strategies & Management. 2002;25(4):726–751. doi:10.1177/1525107118759900.

58. Terrill W, Paoline III EA, Ingram JR. Beyond the final report: A research note on the assessing police use of force policy and outcomes project. Policing: An International Journal. 2018;41(2):194–201. doi:10.1108/PIJPSM-04-2017-0047.

59. Heckman JJ. Sample selection bias as a specification error. Econometrica: Journal of the Econometric Society. 1979; p. 153–161.

60. Buchanan L, Bui Q, Patel JK. Black Lives Matter May Be the Largest Movement in U.S. History. New York Times. 2020;.

61. Silverstein J. The global impact of George Floyd: How Black Lives Matter protests shaped movements around the world. CBS News. 2021;.

62. McGreal C, Beckett L, Laughland O, Ajasa A. Derek Chauvin found guilty of murder of George Floyd. The Guardian. 2021;.

63. Lartey J. Three Years After George Floyd's Murder, Police Reforms Are Slow-Paced. CBS News. 2023;.

64. Freeman JB, Stolier RM, Ingbretsen ZA, Hehman EA. Amygdala responsivity to high-level social information from unseen faces. Journal of Neuroscience. 2014;34(32):10573–10581. doi:10.1523/JNEUROSCI.5063-13.2014.

65. Ericsson KA, Simon HA. Protocol analysis: Verbal reports as data. the MIT Press; 1984.