Social Cognition, Vol. 38, No. 6, 2020, pp. 555–570

WHEN PRACTICE FAILS TO REDUCE RACIAL BIAS IN THE DECISION TO SHOOT: THE CASE OF COGNITIVE LOAD

Balbir Singh
University of Colorado Boulder

Jordan Axt
McGill University

Sean M. Hudson
University of Colorado Boulder

Christopher Lee Mellinger
University of Colorado Boulder

Bernd Wittenbrink
University of Chicago

Joshua Correll
University of Colorado Boulder

Practice improves performance on a first-person shooter task (FPST), increasing accuracy and decreasing racial bias. But rather than simply promoting cognitively efficient processing, we argue that the benefits of practice on a difficult, cognitively demanding task like the FPST rely, at least in part, on resource-intensive, cognitively effortful processing. If practice-based improvements require cognitive resources, then cognitive load should compromise the value of practice by depriving trained participants of the cognitive resources on which they depend. This experiment shows that inducing cognitive load eliminates the benefits of training, leading to an increase in racial bias, as predicted.

Keywords: racial bias, practice, cognitive efficiency, cognitive effort, FPST

Primary support for this work was provided by National Science Foundation Continuing Grant 0642580 to the 6th author. We thank members of the Stereotyping & Prejudice Research Lab at the University of Chicago for helpful comments. Data and supplemental analyses are available at https://osf.io/mwydv/.

Correspondence should be addressed to Balbir Singh, 345 UCB, University of Colorado, Boulder, CO 80309-0345; E-mail: Balbir.singh@colorado.edu

Data from the Department of Justice (2001) suggest that from 1978 to 2001 police killed roughly five times as many Black suspects, per capita, as White suspects. There has been considerable debate as to the reasons why this disparity exists. Some researchers suggest that the disparity stems from racial bias in officers' use of lethal force (Ross, 2015; Scott, Ma, Sadler, & Correll, 2017; Smith, 2004; but see Terrill & Mastrofski, 2002). In particular, the stereotype linking Black men to the concept of danger may lead observers to overestimate the threat posed by Black suspects. This association between Black men and danger is one of the most provocative, commonly studied stereotypes within social psychology. For more than 50 years, researchers have discussed its characteristics and the potential for changing it or at least minimizing its consequences (e.g., Devine, 1989; Dovidio, Evans, & Tyler, 1986; Duncan, 1976; Plant & Devine, 1998; Wittenbrink, Judd, & Park, 1997), including work on the capacity of training to reduce bias in shooting decisions (Correll et al., 2007).

We have investigated the Black-danger stereotype and its implications for racial bias in the decision to shoot using a simple first-person shooter task (FPST), which presents a series of images of Black and White men (Correll, Park, Judd, & Wittenbrink, 2002). Targets appear on screen holding a variety of objects, including cell phones, wallets, and guns. In an effort to capture some of the complexity of an actual police encounter, the targets appear against realistic backgrounds, and the timing and location of their presentation is varied, meaning participants never know when or where to expect a target. The task requires participants to "shoot" armed targets, but to avoid shooting unarmed targets by pressing buttons labeled *shoot* and *don't shoot* as quickly as possible.

In this task, the target's race is formally irrelevant to the decision: participants must simply attend to the object in each target's hand. However, race typically impacts performance. In our initial research (Correll et al., 2002), undergraduates and members of the community showed clear evidence of bias in two separate measures. First, they showed bias in reaction time (RT), shooting armed Black targets more quickly than armed Whites, but choosing *don't shoot* more quickly for unarmed White targets than unarmed Blacks. Second, participants showed bias in the nature of the errors they made. Participants incorrectly shot unarmed Black targets more often than unarmed Whites, and they failed to shoot armed Whites more often than armed Blacks. In other words, whether the target was armed or unarmed, participants were more likely to shoot Black rather than White targets. This bias is apparent when a signal detection theory (SDT) framework is used to analyze the data. Participants employed a more lenient criterion—indicating a greater propensity to shoot—when the target was Black rather than White. Based on these and other data, we argued that participants quickly (if inadvertently) attend to the target's race (Correll, Wittenbrink, Crawford, & Sadler, 2015) and activate task-relevant racial stereotypes concerning danger (Correll et al., 2002, 2007), which create a predisposition to shoot when the target is Black but not when the target is White.

We have also examined bias among police officers—people who make decisions about the use of deadly force in the course of their daily lives. There is reason to expect that police officers might differ from lay people in their performance on tasks like the FPST. Most obviously, police practice making *shoot/don't-shoot* decisions. During their initial training and at regular intervals throughout their careers, officers practice at firing ranges and train with a variety of tools (e.g., interactive video systems) that simulate encounters with suspects. In addition, officers' on-the-job experiences require them to frequently evaluate threat. In encounters with the public, officers must determine whether or not the people around them pose a threat. In many situations, they cannot know for certain whether they are interacting with a harmless individual or an armed criminal. Though the latter case is rare, police must routinely exercise vigilance and assess threat, because failure to do so can have fatal consequences (Department of Justice, 2001, 2006). Routine practice and daily experience in evaluating threat might improve officers' performance on *shoot/don't-shoot* tasks, leading to faster and more accurate decisions and (potentially) reducing the impact of peripheral cues like race.

The results from our research with police officers presented something of a puzzle (Correll et al., 2007). Across multiple studies, we observed two seemingly incongruous effects. First, officer training had *no* observable effect on RT bias. Just like novices, police officers were faster to *shoot* armed Black targets than armed White targets, and were faster to choose a *don't-shoot* response for unarmed White targets than unarmed Black targets. This bias suggests that police (like novices) perceive the target's race and call to mind relevant racial stereotypes. Indeed, when we looked at a subset of officers for whom stereotypes should be particularly salient (i.e., officers from urban areas with large minority populations and high crime rates), we found especially pronounced patterns of RT bias. Based on their RTs, it seems that even police officers were affected by stereotypes associating Black men with danger.[1]

However, a very different pattern emerged for error rates. Unlike novices, who showed clear racial bias in errors, police showed no such bias. More specifically, SDT analyses showed that novices employed a more lenient (or "trigger-happy") criterion for Black than White targets. Police, however, used statistically equivalent criteria for both White and Black targets.[2] Even when we examined officers from high-crime, heavily minority areas, we found no evidence of SDT bias in the decisions they made. Moreover, we observed similar patterns when we gave a group of undergraduates an opportunity to practice the FPST (through completing extra trials). Although the training we give participants is less intense than the training that police officers receive, trained participants looked much like police officers, showing robust RT bias but demonstrating no bias in criteria.

1. We use the term stereotype to refer to an association between race and the concept of danger. Objective differences between Whites and Blacks in the prevalence of criminal activity or danger may give rise to such an association, but the psychological construct (i.e., the stereotype) can be differentiated from actual patterns of covariation between race and danger (see Correll, Park, Judd, & Wittenbrink, 2007, for a discussion of this issue).

2. A number of other effects that are not related to target race suggest that training has benefits: Police officers and trained participants were generally faster and more accurate, and they were overall more conservative in their decisions to shoot.

We wish to highlight the discrepancy between these two patterns of data. Just like novices, police (and trained undergraduates) showed pronounced bias in their RTs; but unlike novices, police (and trained undergraduates) showed no bias in their SDT criteria. The persistent bias in the RT data clearly suggests that police officers and trained participants are not completely immune to stereotypes. But even though these individuals may call stereotypes to mind, and although those stereotypes may influence the *speed* of their responses, stereotypes have no observable influence on whether or not an officer or a trained participant ultimately decides to shoot.

The current work investigates this discrepancy between RT and error measures. Assuming that the RT pattern stems from an underlying, fast-acting, prepotent racial bias, the fact that novices, trained participants, and police exhibit the same RT pattern suggests that practice does not reduce bias by altering the underlying stereotypic associations or through any other sort of routinized, efficient process. Still, practice seems to minimize the impact of stereotypes on the choices people ultimately make.

We suggest that practice reduces bias in error rates at least in part due to relatively inefficient processes. Here, our use of the term *efficiency* is based on Moors & De Houwer (2006), who defined efficiency as a process that "consumes little or no processing resources or attentional capacity" (p. 317). In the following study, we will consider efficient processes to be those that operate despite the presence of other resource demands. For example, a fully routinized or efficient skill, like driving a motor vehicle, will operate smoothly and without error despite the presence of other cognitive demands, such as having a conversation with a passenger. We might say that a driver who requires complete silence to safely control a vehicle has not routinized the skill of driving. They may be able to drive safely under ideal conditions, but if another demand engages processing resources, they will start to make mistakes. To be clear, practice likely routinizes many aspects of performance, but our hypothesis is that at least some of the bias-reducing effects of practice depend on resource-intensive processing.

Our reasoning for this argument is based on the relative complexity of the FPST. Research suggests that task complexity is a critical factor in determining whether practice leads to routinization, on one hand, or cognitively effortful attempts to control behavior, on the other. Practice on simple tasks leads to routinization; practice on more complex tasks can promote effortful attempts to exert control (Olesen, Westerberg, & Klingberg, 2004; Shiffrin & Schneider, 1977). Again, note that the FPST is considerably more complicated than typical social cognitive tasks, which often use simple stimuli such as a cropped face, an isolated object, or a single word, and which present these stimuli against a simple, uniform background (e.g., Correll, 2011; Donders, Correll & Wittenbrink, 2008; Fazio, Jackson, Dunton, & Williams, 1995; Ito & Urland, 2003; Payne, 2001; Wittenbrink et al., 1997). The FPST employs full-body images of 50 different targets situated in 20 different background scenes, each of which is a complex, realistic photograph. The task also varies the timing and location of target presentation, so that participants never know where the target will appear or when they will have to react. Once a target appears, participants must determine whether a small object in the target's hand is a gun or something harmless, like a cell phone. In the FPST stimuli, the object typically accounts for roughly 1/500th or 0.2% of the image area. On such a complex task, performance may benefit, in part, from the development of cognitively intensive control-related processes (Shiffrin & Schneider, 1977). These effects of training may accrue after extended practice with the FPST, itself, but they may also accrue from day-to-day police work. It thus seems plausible that repeatedly exercising vigilance for threats in real-world situations reduces bias in this simulation in part because it promotes control (Correll et al., 2007).

If training reduces SDT bias in part because trained participants are exerting effort to control their performance, it stands to reason that SDT bias should reemerge when trained participants are deprived of executive resources. On the other hand, if training simply promotes efficient, routinized processing, cognitive load should not disrupt trained performance. The current experiment tests the effect of concurrent cognitive load on novices and trained participants (whose prior performance mirrors that of police in critical ways). We expose participants to different levels of cognitive load during the FPST and examine the consequences of this manipulation on task performance. We predict that, although trained participants may show reduced SDT bias in the absence of a cognitive load (when they can devote their full resources to controlling performance on the task), they should show increased levels of bias when cognitive resources are diverted by a secondary task.

In one sense, the idea that control and cognitive load play a critical role in FPST performance is no surprise. As an individual difference, more effective control (or executive function) is associated with reductions in bias in tasks like the FPST (Ito et al., 2015; Payne, 2005). Further, as noted above, practice on these tasks typically improves accuracy (i.e., increasing sensitivity in SDT or control-related estimates in process dissociation (PDP), Correll et al., 2007; Plant, Peruche, & Butz, 2005). Finally, manipulations that deprive participants of cognitive resources typically reduce overall performance accuracy (Govorun & Payne, 2006; Kleider, Parrott, & King, 2009; Payne, 2001).

So, how does this work add to the existing literature? The current study offers two novel and important contributions. First, although previous work does indeed show that practice improves performance accuracy, which can be measured using PDP's "control" estimate, there is currently no evidence that the improvement actually relies on *controlled cognitive operations*. Through practice, a participant may learn cognitive shortcuts that improve accuracy through simple and cognitively efficient responses (e.g., images of guns may all have a slight blue cast). Accordingly, the PDP control parameter might reflect increases in accuracy that are due to mental operations that would typically be considered more automatic than controlled. Second, manipulations that deprive participants of cognitive resources, like fatigue or cognitive load, have generally been shown to reduce accuracy (e.g., SDT sensitivity or PDP control), but they have not been shown to increase racial bias (for the lone exception, see Ma et al., 2013). In the current study, we deviate from all previous work by testing the effect of cognitive load on trained participants who typically show no SDT bias. Again, we have reason to believe that those practice-based reductions in bias rely partly on controlled operations. We can therefore predict that induction of a cognitive load will not simply reduce accuracy, but will actually reintroduce racial bias. This study will test and, ultimately, provide evidence for both of these novel and important predictions.

METHOD

PARTICIPANTS AND DESIGN

One hundred thirty-nine non-Black undergraduates (68 White, 52 Asian; 16 Latino, 2 Arabic, 1 who did not report race/ethnicity; 76 female) at the University of Chicago were randomly assigned to either a trained or novice condition. Trained participants practiced the standard FPST, while novices did not. Both groups then performed the FPST under three levels of cognitive load (low, medium, and high, manipulated within-participant), which was induced by presenting auditory stimuli and asking them to make judgments of varying difficulty. In each load condition, participants completed 50 trials of the FPST, which involved White and Black targets paired with guns and harmless objects. Accordingly, the study involved a 2 (Training: novice vs. trained) × 3 (Cognitive Load: low vs. medium vs. high) × 2 (Target Race: Black vs. White) × 2 (Object Type: armed vs. unarmed) mixed-model design, with repeated measures on the last 3 factors. We began recruitment in the fall and continued till the end of the academic year, recruiting as many participants as we could. Analyses were run after data collection finished.

MATERIALS

The goal of this study was to replicate previously documented effects of practice as a way to reduce racial bias, but then to undermine the benefit of that practice with a cognitive load. Because previous work only found practice effects with respect to bias in error rates, this study looked to increase variability of error rates. As in past research, this FPST used a short time window (see Correll, 2011; Correll et al., 2002, 2007, 2015), requiring participants to respond within 630 ms. Responses after 630 ms were not collected. Though advantageous for testing variability in error rates, this window constrains variability in RTs (Correll et al., 2002), so we focus exclusively on the error rates, using the lens of SDT.

The standard FPST was modified to present an auditory cue once per trial. These auditory cues consisted of digital recordings of a male voice naming single digit numbers (0–4 and 6–9). The presentation of these numbers varied, such that the number was randomly presented between 0 and 600 ms prior to the FPST target image. The FPST included three 50-trial blocks, which varied only in the nature of the participant's instructions. In the low-load block, participants were instructed simply to perform the FPST and ignore the auditory cue. In the medium-load block, in addition to performing the FPST, participants were instructed to indicate during the trial whether the number named in the auditory cue was greater or less than

5 by saying either "high" or "low." In the high-load block, they were instructed to indicate whether the number presented on the current trial was greater or less than the number presented on the previous trial by saying either "high" or "low." Thus, the high-load condition forced participants to hold one number in memory and compare the present cue to that stored value (a 1-back task). This manipulation has been shown to tax working memory (Conway et al., 2005; Kirchner, 1958). In the medium- and high-load blocks, the experimenter recorded participants' responses (high, low, or no response) for each trial.

Because experience with the task (i.e., practice) can affect performance, we counterbalanced the order of the three blocks. Participants were randomly assigned to complete the blocks in one of six possible orders. Though block order is not of interest in the present study, we controlled for its effects.[3]

PROCEDURE

Participants were greeted by either a White female or male researcher. They were asked to complete a consent form and were randomly assigned to either the novice or trained conditions. Participants in the trained condition completed 116 trials of the FPST before starting the test phase. During this process, they were not required to respond to auditory cues. Novices proceeded directly to the test phase after completing the consent form and 16 practice trials. All participants then completed the 3-block version of the FPST, which manipulated cognitive load (low load vs. medium load vs. high load) within participants. We collected demographic information, and participants were fully debriefed. No other measures were collected, and no other manipulations occurred.

RESULTS AND DISCUSSION

EXCLUSION CRITERIA

To meaningfully test the study's central hypothesis, we needed to ensure that participants were reasonably well engaged in both the FPST and the auditory task. Individuals who either ignored the FPST (to focus exclusively on the auditory task) or who ignored the auditory task (to focus exclusively on the FPST) would not provide meaningful data (e.g., Turner & Engle, 1989). We therefore excluded participants whose error rate on the auditory task exceeded 20% (3 novices, 4 trained) or failed to respond within the time window on more than 30% of the FPST trials in either the medium- or the high-load condition (6 novices, 3 trained).[4]

3. The inclusion of Block Order in our analysis does not impact the pattern of results. We would like to note that block order interacted with the linear effect of cognitive load on d'. This interaction was not systematic and defied any obvious interpretation. It was not, for example, a product of completing the high- versus low-load block at the beginning of the study. Thus, we attribute this significant result to a Type I error. In the supplemental analysis we include every model that we ran, and further decompose the significant interaction of the linear effect and block order on d'.

4. The primary three-way interaction reported below remains significant for cutoffs ranging from 10–40%.

We also excluded one novice who timed out on every high-load trial involving an unarmed White target, making it impossible to compute SDT estimates. Thus, we had a total of 122 participants in the analyses. With this sample, a sensitivity analysis indicates we had an 80% chance of detecting an effect equivalent to $\eta^2 = 0.064$.

SIGNAL DETECTION ANALYSIS[5]

We computed SDT estimates of sensitivity (*d′*) and decision criterion (*c*) separately for White and Black targets within each of the three levels of cognitive load. These estimates, which are summarized in Table 1, were submitted to a 2 (Training) × 3 (Cognitive Load) × 2 (Target Race) mixed-model analysis, focusing on the linear effects of Cognitive Load.

Sensitivity. To ensure that the cognitive load manipulation was effective, we examined participants' ability to perform the FPST. In particular, we examined SDT estimates of sensitivity, which represent the participants' ability to distinguish between armed and unarmed targets and respond appropriately. In line with prior findings (Correll et al., 2007), we observed a marginal effect of training, such that trained participants showed higher sensitivity than novices, $b = 0.145$, $F(1,115) = 3.21$, $p = .076$, $\eta^2 = 0.027$, 95% CI [–0.015, 0.306]. Further, cognitive load impaired performance. As load increased (the linear effect), sensitivity decreased, $b = –1.355$, $F(1,115) = 49.57$, $p < .001$, $\eta^2 = 0.301$, 95% CI [–1.737, –0.974]; see top panel of Figure 1. We also observed a quadratic effect of Load, $b = –1.118$, $F(1,115) = 11.76$, $p = .001$, $\eta^2 = 0.093$, 95% CI [–1.764, –0.472]: there was a decrease in sensitivity from the low-load to the medium-load conditions, $b = –1.237$, $F(1,115) = 42.65$, $p < .001$, $\eta^2 = 0.271$, 95% CI [–1.612, –0.862] , whereas sensitivity in the medium- and high-load conditions was statistically indistinguishable, $b = –0.118$, $F(1,115) = 0.39$, $p = 0.533$, $\eta^2 = 0.003$, 95% CI [–0.494, 0.257]. The main effects of Training and Load suggest that our manipulations were effective. An unanticipated main effect of Target Race emerged, $b = –0.947$, $F(1,115) = 17.62$, $p < .001$, $\eta^2 = 0.133$, 95% CI [–1.393, –0.500], such that sensitivity was higher for Black targets than for White targets.

Criteria. Our primary question involved the magnitude of racial bias in the SDT criteria. We found clear evidence of bias, manifested as a main effect of Target Race, $b = 0.527$, $F(1,115) = 20.37$, $p < .001$, $\eta^2 = 0.151$, 95% CI [0.295, 0.757], such that participants set a lower criterion (greater tendency to choose *shoot*) for Black targets than White targets. Replicating previous data with trained participants and police

5. Based on reviewer suggestions, we also analyzed these results through Process Dissociation Procedure (PDP; Jacoby, 1991) and through a Bayesian hierarchical drift diffusion model (HDDM) framework (Johnson et al., 2017; Ratliff & McKoon, 2008). The PDP analysis yielded the same pattern of results with corresponding PDP and SDT estimates correlating above r = .89. We chose to report the SDT analysis, as the FPST conforms to curvilinear ROC curves (Rotello, Kelly, & Heit, 2018), an assumption of SDT (Macmillan & Creelman, 2005). The HDDM analysis did not yield any novel insights as the relatively small trial numbers in the current dataset reduce the precision with which HDDM parameters can be estimated, limiting power to detect meaningful effects. For any interested reader, we have made both the PDP analysis and the HDDM analysis available at https://osf.io/mwydv/.

TABLE 1. Means (and Standard Deviations) of Behavioral and SDT Estimates of Sensitivity and Criterion (Lower Scores Indicate Greater Tendency to Shoot) as a Function of Training, Target Race and Cognitive Load

	Novice (n = 68)					
	Low Cognitive Load		Medium Cognitive Load		High Cognitive Load	
	Black	White	Black	White	Black	White
Error Rates	0.15 (0.36)	0.17 (0.38)	0.22 (0.42)	0.25 (0.43)	0.21 (0.41)	0.28 (0.45)
Hits	0.89 (0.13)	0.82 (.14)	0.80 (0.16)	0.75 (0.19)	0.81 (0.18)	0.70 (0.22)
False Alarms	0.20 (0.17)	0.17 (0.16)	0.27 (0.25)	0.26 (0.23)	0.26 (0.21)	0.28 (0.20)
Criterion	–0.25 (0.53)	0.09 (0.48)	–0.12 (0.71)	–0.01 (0.71)	–0.12 (0.60)	0.05 (0.53)
d Prime	2.75 (1.35)	2.30 (1.20)	1.89 (1.31)	1.70 (1.23)	2.10 (1.43)	1.43 (1.34)
Reaction Time (Armed)	495.80 (63.81)	505.14 (63.14)	497.66 (70.35)	503.49 (67.88)	506.08 (67.71)	501.37 (72.15)
Reaction Time (Unarmed)	531.95 (62.88)	533.67 (62.63)	527.39 (68.21)	523.73 (71.02)	526.18 (73.49)	527.71 (70.85)
	Trained (n = 54)					
	Low Cognitive Load		Medium Cognitive Load		High Cognitive Load	
	Black	White	Black	White	Black	White
Error Rates	0.15 (0.36)	0.16 (0.36)	0.19 (0.39)	0.21 (0.41)	0.20 (0.40)	0.23 (0.42)
Hits	0.86 (0.11)	0.85 (0.12)	0.83 (0.12)	0.78 (0.15)	0.81 (0.16)	0.74 (0.16)
False Alarms	0.16 (0.15)	0.17 (0.17)	0.21 (0.19)	0.21 (0.19)	0.22 (0.16)	0.19 (0.15)
Criterion	–0.03 (0.47)	–0.02 (0.60)	–0.01 (0.56)	0.12 (0.50)	–0.06 (0.51)	0.17 (0.40)
d Prime	2.61 (1.24)	2.52 (1.07)	2.22 (1.33)	1.97 (1.11)	2.15 (1.32)	1.96 (1.32)
Reaction Time (Armed)	482.82 (64.46)	491.50 (66.60)	489.75 (67.99)	488.83 (69.79)	493.66 (66.55)	500.07 (71.04)
Reaction Time (Unarmed)	523.47 (61.50)	520.62 (59.77)	519.84 (66.72)	519.52 (67.67)	521.76 (63.58)	526.20 (62.85)

officers, we observed a marginal main effect of Training, $b = 0.051$, $F(1,115) = 3.32$, $p = .071$, $\eta^2 = 0.028$, 95% CI [–0.004, 0.106], such that trained participants generally used higher, more conservative criteria (Correll et al, 2007). That is, trained participants choose to *shoot* less often across all trial types. None of the two-way interactions approached significance.

Most importantly, the three-way Training × Cognitive Load (linear) × Target Race interaction, which is central to the current research hypothesis, was significant, $b = 0.206$, $F(1,115) = 6.58$, $p = .012$, $\eta^2 = 0.054$, 95% CI [0.047, 0.366], see Figure 1, bottom panel. To clarify this interaction, we examined the effect of Training on bias in each of the three load conditions. In the low-load condition, with minimum distraction, we replicated previous work showing that Training reduces racial bias in SDT criteria, $b = –0.171$, $F(1,115) = 7.06$, $p = .009$, $\eta^2 = 0.058$, 95% CI [–0.299, –0.043]. Compared to novices, trained participants showed a reduction in bias. But as cognitive load increased from low to medium to high, the benefit of practice disappeared (this linear change is tested by the three-way interaction, reported above).

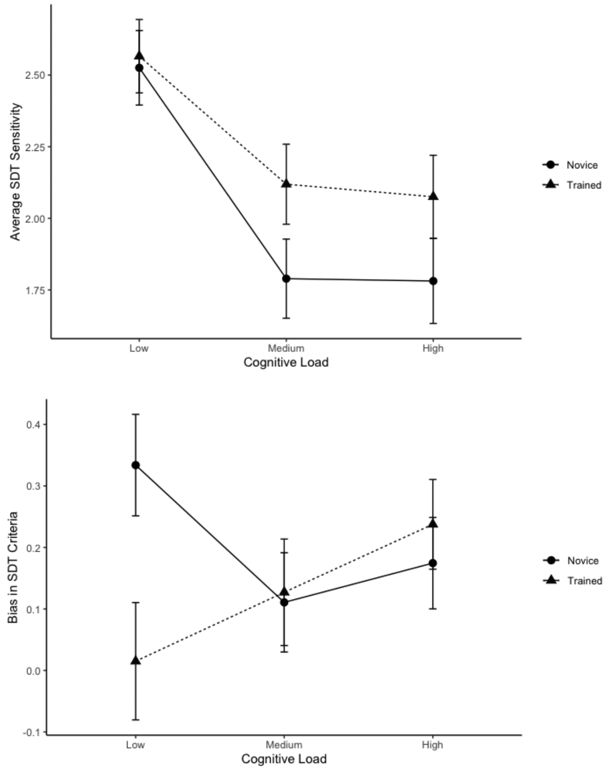

FIGURE 1. Mean SDT sensitivity (*d′*, top panel) and racial bias (*c*, bottom panel; higher numbers indicate a greater tendency to shoot Black compared to White targets) as a function of Training and Cognitive Load.

In the medium-load condition, the difference between trained participants and novices was no longer significant, $b = 0.019$, $F(1,115) = 0.10$, $p = 0.751$, $\eta^2 = 0.001$, 95% CI [–0.101, 0.139]. Similarly, when Cognitive Load was highest, trained participants did not differ statistically from novices, $b = 0.035$, $F(1,115) = 0.43$, $p = 0.511$, $\eta^2 = 0.004$, 95% CI [–0.071, 0.141]. Moreover, these trained participants now showed statistically significant racial bias themselves, using a significantly lower criterion for Black targets than for White targets, $b = 0.254$, $F(1,115) = 10.41$, $p = .002$, $\eta^2 = 0.083$, 95% CI [0.098, 0.411], (directionally, they showed *more* bias than novices). This is a critical point: When subjected to a cognitive load, the trained participants showed as much racial bias in response criterion as novices. The benefits of practice were completely erased.

We can also consider the effect of increasing load separately for trained participants and for novices. Again, if trained participants rely on cognitive control to overcome racial bias in the decision to shoot, cognitive load should exacerbate bias by depriving them of resources they would otherwise use to control their performance. In line with this argument, trained participants showed more bias as cognitive load increased (the linear effect), $b = 0.238$, $F(1,115) = 4.00$, $p = .048$, $\eta^2 = 0.034$, 95% CI [0.002, 0.474]. For novices, by contrast, higher load had no effect on bias—if anything, higher load reduced bias, $b = -0.174$, $F(1,115) = 2.615$, $p = .109$, $\eta^2 = 0.022$, 95% CI [–0.388, 0.039]. This latter trend was not significant and was not predicted a priori. It may reflect the fact that novices were sufficiently overwhelmed by the difficulty of the high-load task that they actually failed to process racial information (e.g., Gilbert & Hixon, 1991). Given the lack of a clear theoretically based prediction for this pattern and non-significant effect, we hesitate to draw any conclusions. No other effects were significant.

GENERAL DISCUSSION

This experiment was a first step in solving a puzzle with both applied and theoretical implications. Previous research on the decision to shoot suggests that police officers show pronounced racial bias in their RTs. For example, they shoot armed Black targets more quickly than armed White targets. But these same officers do *not* show bias in the errors they commit. For example, they do not shoot more unarmed Black targets than unarmed White targets. More specifically, SDT analyses show that officers use statistically equivalent criteria when evaluating both Black and White targets. This dissociation between RT bias and SDT bias suggests that although police officers may activate racial stereotypes about Blacks and danger, they somehow override them when making their decision about whether or not to shoot. In our laboratory simulation and even more so in the real world, the evaluation of a threatening suspect is a multifaceted, complex, and cognitively difficult task. Accordingly, extensive practice may not lead simply to automation of all components of the decision process. Instead, practice may enhance control over some aspects, and this control may be critical for reducing bias.

To explore this possibility, we examined the degree to which practice-based reductions in SDT bias rely on resource-intensive cognitive processes. We accomplished this by manipulating cognitive load. In the absence of load, we replicated previous effects: trained participants were able to avoid showing bias. But when a cognitive load deprived them of executive resources, SDT bias reemerged. This pattern suggests that the benefits of practice on the FPST rely, at least in part, on cognitively effortful processing.

At first glance, these effects may seem to run contrary to the literature on practice and expertise (which focuses primarily on the idea that practice promotes "automaticity"). However, there is substantial support for the idea that, when a judgment task is too complex or difficult to routinize, practice promotes effortful cognitive processing (Ackerman, 1987, 1988; Shiffrin & Schneider, 1977). In their seminal work, Shiffrin and Schneider (1977) found that practice promoted routinization only when the judgment task could be simplified. When the same basic task precluded simplification (e.g., by using variable stimulus-response mapping), participants with extensive practice showed clear evidence of an increase in effortful processing. More recently, studies using brain imaging have found when a task or stimulus set is difficult or complex, practice can promote more extensive activation in areas associated with executive functioning, such as the dorsolateral prefrontal cortex (Miotto, et al., 2006; Moore, Cohen, & Ranganath, 2006; Olesen et al., 2004; Weissman, Woldorff, Hazlett, & Mangun, 2002). These findings parallel the current results.

There are at least two component processes that might underlie practice effects on the FPST. First, police officers and trained participants might learn to directly counteract active stereotypes (e.g., Wegener & Petty, 1995). If an officer knows that Black targets conjure up associations with danger, then, when confronted with a Black target, the officer may adjust his or her expectations in a counterstereotypic direction. This process would constitute a *race-based* solution to a race-based bias, in that it is the target's race (the same cue that ostensibly activates stereotypes) that provides the cue for compensatory adjustment. To our knowledge, only one race-based strategy has been shown as effective (Stewart & Payne, 2008). We view this process as relatively implausible due to evidence that strategic attempts to avoid using race on tasks like the FPST often backfire (Correll, 2008; Payne, Lambert, & Jacoby, 2002).

A more likely possibility is that, through practice, police officers and trained participants learn to more effectively identify and use information *other than race.* In particular, it is possible that police officers and trained participants learn to more effectively attend to whatever object is in the hands of the target (Is it a gun? Is it a cell phone?), interpret that information, and allow it to guide their decision. In line with this possibility, trained participants under high load are relatively more accurate than novice participants, despite both being undermined by cognitive load. Even if we assume that stereotypes induce a stronger tendency to shoot when the target is Black rather than White, the ability to quickly and accurately identify the object may enable a police officer or a trained participant to make an unbiased decision about shooting. Thus, practice effects may stem from these individuals' ability to identify and respond to the object. A prepotent racial bias may still exist (resulting in bias in RTs), but police and trained participants may be better at adjusting their responses based on the object (reducing error rate bias). Such a process would represent an *object-based* adjustment to race-based bias (e.g., Correll et al., 2015). Critically, to the extent that it is difficult to attend to, interpret, and/or utilize visual information about the object, the bias-reducing benefits of training may require cognitive resources. Accordingly, when trained participants are deprived of those resources, bias re-emerges.

If training promotes resource-intensive processing, one might expect to observe that trained participants are slower than novices, particularly under load. We see no evidence of such an effect. Though we can only speculate, we suggest that practice probably routinizes some (simpler) elements of a decision, even if it promotes more elaborate processing of other (more difficult) elements. The current study was not designed to isolate component processes involved in the FPST, but the data suggest that processes were impacted by practice in different ways. Specifically, with reduced cognitive resources, trained participants no longer show a reduction in bias, but they continue to discriminate between armed and unarmed targets more accurately than untrained participants. Practice may thus have routinized some components of performance, which improved SDT sensitivity and/or freed other cognitive resources. The identification and exploration of individual component processes that underlie judgments and behaviors represent challenges for future research.

From a more applied perspective, the current findings are important because they highlight a potential problem for police officers who want to avoid using racial stereotypes when making decisions in the field. In optimal conditions (e.g., the low-load condition in the current study), the data suggest that practice promotes unbiased judgments. Even if racial stereotypes become activated, trained officers (unlike lay people) may be able to formulate decisions based on more relevant information (such as the behavior of the suspect). However, analogous to the high-load condition in this experiment, police do not always have the luxury of operating in optimal conditions. They are often tired and overworked, and may experience extreme arousal when they believe a suspect poses a lethal threat. Fatigue and intense emotions such as fear may compromise executive function, and so affect an officer's ability to implement unbiased decisions (Kleider et al., 2009; Ma et al., 2013; Pessoa, 2009; van der Linden, Frese, & Meijman, 2003). It is important to note that many police shootings occur late at night (implicating fatigue) and in dangerous neighborhoods (implicating fear) (Vila, 1996; Villa & Kenney, 2002). If training reduces bias only in optimal test conditions, it may not be adequate. This raises the disturbing possibility that discrepancies in actual police shootings (Ross, 2015) reflect the fact that police must make these decisions in challenging, exhausting, and frightening conditions that tax their cognitive capacity, nullifying any prior training that could have improved their ability to override or ignore racial stereotypes.

This conclusion seems fairly bleak, and we recognize that it seems to challenge the value of practice and training in the first place. However, we know that practice will generally be most effective when it mirrors test conditions, a principle known as structural reinstatement (for a review, see Healy, Wohldmann, & Bourne, 2005). Police departments nationwide have made tremendous changes to their training over the last century reflecting this basic idea (Fyfe, 1981; Geller, 1982; Grossman, 2004). In one stunning example, Grossman observed that, in the past, when police trained at the firing range, officers were instructed to empty the spent shell casings from their revolvers into their hands, then put the casings in their pockets. The goal of this practice was to save on janitorial costs. But officers who were killed in gun battles were later found with their pockets full of shell casings. In the midst of a life-or-death situation, rather than quickly and efficiently dropping the casings on the street and reloading, they had taken the time to diligently pocket the spent brass. In the panic of a fight, they fell back on the habits they had practiced.

The simple lesson is that officers must prepare for the conditions they face in the field, where being fast is more valuable than being tidy. In the same way, if we know that a typical police encounter involves fear and fatigue, it may be valuable for officers to practice making decisions when they are scared and tired. Experience dealing with an emotionally and cognitively demanding task may help officers deal with similar situations in the street. By way of example, Beilock and

Carr (2001) induced high-pressure situations for golf players, causing even highly trained individuals to "choke" when putting. But participants who had trained in a high-pressure environment were less likely to choke. Their training had prepared them not only make the putt, but to make the putt while simultaneously dealing with the relevant anxiety.

Police departments have multiple options for training their officers, some more realistic than others. Firing ranges may emphasize speed and accuracy, but there is little external validity. Video-based simulators, in which the trainee "interacts" with a dynamic (if two-dimensional) suspect, can create more psychologically engaging and cognitively complex environments. Some departments even have the resources to create a Hogan's alley, a fake city street like a movie set, where the trainee interacts with actors who are armed with guns that fire nonlethal (but painful) ammunition. Similar intensive training environments are used in the military, where for instance, Navy Seals go through "Hell Week," surviving sleep deprivation, undergoing tests of physical strength, enduring extreme cold, and so forth. Though further work is necessary to examine the effects of different training regimens, realism may provide officers with an opportunity to practice handling the full cognitive, physical, and emotional complexity of a potentially lethal encounter.

REFERENCES

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3–27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288–318.

Beilock, S. L. & Carr. T.H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130(4),* 701–725.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769–786.

Correll, J. (2008). 1/f noise and effort on implicit measures of bias. *Journal of Personality and Social Psychology*, *94*(1), 48–59.

Correll, J. (2011). Order from chaos? 1/f noise predicts performance on reaction time measures. *Journal of Experimental Social Psychology*, *47*(4), 830–835.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83,* 1314–1329.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92,* 1006–1023.

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: How stereotypes disambiguate visual stimuli. *Journal of Personality and Social Psychology*, *108*(2), 219–233.

Department of Justice. (2001). *Policing and homicide, 1976–98: Justifiable homicide by police, police officers murdered by felons* (NCJ 180987). Washington, DC: Bureau of Justice Statistics.

Department of Justice. (2006). *Violent encounters: A study of felonious assaults on our nation's law enforcement officers.* Clarksburg, WV: Federal Bureau of Investigation Uniform Crime Reports.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Donders, N.C., Correll, J., & Wittenbrink, B. (2008). Danger stereotypes predict racially biased attentional allocation. *Journal of Experimental Social Psychology. 44(5),* 1328–1333.

Dovidio, J., Evans, N., & Tyler, R. (1986). Racial stereotypes: The contents of their cognitive representation. *Journal of Experimental Social Psychology, 22*, 22–37.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology*, *34*(4), 590–598.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Fyfe, J.J. (1981). Race and extreme police-citizen violence. In R. McNeely and C. Pope (Eds.), *Race, crime, and criminal justice* (pp. 89–108). Beverly Hills, CA: Sage.

Geller, W. A. (1982). Deadly force: What we know. *Journal of Police Science and Administration, 10*, 151–177.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.

Govorun, O., & Payne, B. K. (2006). Ego—Depletion and prejudice: Separating automatic and controlled components. *Social Cognition*, *24*(2), 111–136.

Grossman, D. (2004). *On combat*. Bellville, IL: PPCT Research.

Healy, A. F., Wohldmann, E. L., & Bourne, L. E. (2005). The procedural reinstatement principle: Studies on training, retention, and transfer. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 59–71). Washington, DC: American Psychological Association.

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, *108*(2), 187–218.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*, 616–626.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358.

Kleider, H. M., Parrott, D. J., & King, T. Z. (2009). Shooting behavior: How working memory and negative emotionality influence police officer shoot decisions. *Applied Cognitive Psychology*, *24*(5), 707–717.

Ma, D. S., Correll, J., Wittenbrink, B., Bar-Anan, Y., Sriram, N., & Nosek, B. A. (2013). When fatigue turns deadly: The association between fatigue and racial bias in the decision to shoot. *Basic and Applied Social Psychology*, *35*(6), 515–524.

Miotto, E. C., Savage, G. R., Evans, J. J., Wilson, B. A., Martins, M. G. M., Laki, S., & Amaro, E. (2006). Bilateral activation of the prefrontal cortex after strategic semantic cognitive training. *Human Brain Mapping, 27(4),* 288–295.

Moore, C. D., Cohen, M.X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *Journal of Neuroscience, 26*(43), 11187–11196.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297–326.

Olesen P., Westerberg H., & Klingberg T. (2004). Increased prefrontal and parietal brain activity after training of working memory. *Nature Neuroscience, 7*(1), 75–79.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality & Social Psychology, 81*, 1–12.

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, *89*(4), 488–503.

Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38*, 384–396.

Pessoa, L. (2009). How do emotion and motivation direct executive control. *Trends in Cognitive Sciences, 13*, 160–166.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75,* 811–832.

Plant, E. A., Peruche, B. M., & Butz, D. A. (2005). Eliminating automatic racial bias: Making race non-diagnostic for responses to criminal suspects. *Journal of Experimental Social Psychology*, *41*(2), 141–156.

Ross, C. T. (2015). A multi-level Bayesian analysis of racial bias in police shootings at the county-level in the United States, 2011–2014. *Plos One, 10*(11).

Scott, K., Ma, D. S., Sadler, M. S., & Correll, J. (2017). A social scientific approach toward understanding racial disparities in police shooting: Data from the Department of Justice (1980–2000). *Journal of Social Issues, 73*(4), 701–722.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Smith, B. W. (2004). Structural and organizational predictors of homicide by police. *Policing: An International Journal of Police Strategies and Management, 27,* 539–557.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin, 34*(10), 1332–1345.

Terrill, W., & Mastrofski, S.D. (2002). Situational and officer-based determinants of police coercion. *Justice Quarterly, 19*(2), 215–248.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica, 113*, 45–65.

Vila, B. (1996). Tired cops: Probable connections between fatigue and the performance, health, and safety of patrol officer. *American Journal of Police*, *2,* 51–92.

Vila, B., & Kenney, D. J. (2002). Tired cops: The prevalence and potential consequences of fatigue. *National Institute of Justice Journal*, *248*, 17–31.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, *68*(1), 36–51.

Weissman, D. H., Woldorff, M. G., Hazlett, C. J., & Mangun, G. R. (2002). Effects of practice on executive control investigated with fMRI. *Cognitive Brain Research, 15,* 47–60.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274.