Title
Emotion expression salience and racially biased weapon identification: A diffusion modeling approach

Authors
Klein, Samuel AW
Todd, Andrew R

Publication Date
2024-05-20

DOI
10.3758/s13423-024-02526-z

Emotion Expression Salience and Racially Biased Weapon Identification:

A Diffusion Modeling Approach

Samuel A. W. Klein[1*] and Andrew R. Todd[1*]

[1] Department of Psychology, University of California, Davis

This is a preprint of a manuscript that was published in Psychonomic Bulletin & Review

[*]Corresponding authors:

E-mail: sawklein@ucdavis.edu or atodd@ucdavis.edu

Phone: +1 (530) 752-1880

This research was facilitated by NSF grant BCS-1764097 (awarded to ART).

Abstract

Racial stereotypes are commonly activated by informational cues that are detectable in people's faces. Here, we used a sequential priming task to examine whether and how the salience of emotion (angry/scowling vs. happy/smiling expressions) or apparent race (Black vs. White) information in male face primes shapes racially biased weapon identification (gun vs. tool) decisions. In two experiments (N_{total} = 546) using two different manipulations of facial information salience, racial bias in weapon identification was weaker when the salience of emotion expression versus race was heightened. Using diffusion modeling, we tested competing accounts of the cognitive mechanism by which the salience of facial information moderates this behavioral effect. Consistent support emerged for an initial bias account, whereby the decision process began closer to the "gun" response upon seeing faces of Black versus White men, and this racially biased shift in the starting position was weaker when emotion versus race information was salient. We discuss these results vis-à-vis prior empirical and theoretical work on how facial information salience moderates racial bias in decision-making.

Emotion Expression Salience and Racially Biased Weapon Identification:

A Diffusion Modeling Approach

Racial stereotypes pervade modern thinking, with abundant experimental evidence more strongly linking Black versus White people with weapons (Payne & Correll, 2020). In the weapon identification task (WIT), for example, participants are usually better (i.e., faster and more accurate) at identifying guns and worse at identifying harmless objects (e.g., tools, toys) after seeing Black versus White face primes (Amodio et al., 2004; Payne, 2001; Todd et al., 2016). This typical pattern of racial bias in the WIT is robust (see Rivers, 2017); however, its magnitude may vary by the salience of (i.e., the attention garnered by; Higgins, 1996) information in the face primes. Indeed, racially biased weapon identification is weaker and sometimes eliminated when age versus race information is more salient (Jones & Fazio, 2010; Todd et al., 2021). Granted, age is only one of many sources of social information. In two experiments, we investigated whether attending to another facial cue—emotion expression—likewise weakens weapon-related racial bias, relative to attending to race.

Unlike facial cues pertaining to relatively static social categories (e.g., age, race), emotion expressions can vary moment-to-moment within the same target person. Furthermore, emotion expressions presumably signal affect and intentions (Niedenthal & Brauer, 2012; Todorov et al., 2008) in ways that demographic cues may not, making them informative for basic social judgment (e.g., identifying threats). Accordingly, emotion expressions may garner substantial attention in threat-related contexts like weapon identification, effectively competing against the attention often garnered by race in such contexts (Payne & Correll, 2020). Indeed, the mere availability of scowls and smiles has been found to affect racially biased weapon identification (Kubota & Ito, 2014), whereas the mere availability of other information (e.g., age cues) has not (Todd et al., 2016). Thus, it seems reasonable to propose that directing attention toward emotion versus race cues moderates racially biased weapon identification. Our experiments tested this proposition.

Besides investigating *whether* the salience of emotion versus race information alters racial bias in the WIT, we examine *how* such an effect emerges using diffusion decision modeling (DDM; Ratcliff et al., 2016). The DDM is a sequential sampling technique designed to disentangle processes underlying behavior in tasks like the WIT. By concurrently modeling both decisions and decision speed, the DDM decomposes decisions into four parameters (see Table 1). We briefly describe two relevant parameters that might explain how information salience moderates racially biased weapon identification.

Table 1

Parameters of the Diffusion Decision Model in the Weapon Identification Task

Parameter	Interpretation
Relative start point (β)	Initial bias to select *gun* or *tool* at the start of evidence accumulation, with $0 < \beta < 1$. Values >.50 indicate a bias to select *gun*; values <.50 indicate a bias to select *tool*.
Threshold separation (α)	Amount of evidence required to decide, with $0 < \alpha$. Hitting a threshold triggers a decision to select *gun* or *tool*.
Drift rate (δ)	Average quality of information extracted at each unit of time, with $-\infty < \delta < \infty$. Higher absolute values indicate stronger evidence. Positive values indicate evidence to select *gun*; negative values indicate evidence to select *tool*.
Non-decision time (τ)	Length of all response components (encoding time, motor response time, and other unknown contaminants) unrelated to decision making, with $0 < \tau$. Measured in milliseconds.

The DDM assumes that evidence is accumulated over time until a decision threshold is reached. It models both the strength of evidence extracted (i.e., drift rate) and the position from which evidence accumulation begins (i.e., relative start point; see Figure S13). An *evidence accumulation* account of facial information salience moderating racially biased weapon identification posits that seeing a Black versus White face prime strengthens the evidence accumulated for identifying guns (i.e., race-stereotypic objects), but that racially biased evidence accumulation is weaker when emotion versus race information is salient. Alternatively, an *initial bias* account posits that seeing a Black versus White face prime shifts the starting position of the decision process closer to the "gun" response, but that shifts in the start point are less racially biased when emotion versus race is salient.[1]

The initial bias account is more strongly supported in the WIT literature.[2] Specifically, whereas an evidence accumulation account did not explain racially biased weapon identification, or its moderation by the salience of age versus race information in the face primes, an initial bias account did (Todd et al., 2021). Relative to age cues, the arguably greater relevance of emotion expression in threat-related contexts might undermine racially biased weapon identification by altering the interpretation of object-related content—the process-level pattern predicted by an evidence accumulation account. Thus, testing these accounts when emotion versus race salience varies in the context of weapon identification is instructive. Our experiments provide such a test.

[1] We did not derive clear predictions about threshold separation and non-decision time, but we report results pertaining to these parameters for completeness.

[2] Notably, an evidence accumulation account better explains racial bias in the first-person shooter task (FPST, Correll et al., 2015; Johnson et al., 2018; Pleskac et al., 2018). For a discussion of procedural differences between the FPST and the WIT, see Todd et al. (2021).

For consistency with prior work (e.g., Todd et al., 2021), we report behavioral analyses of the error rates and correct response times (RTs) along with our analyses of the DDM parameters. Data and code are available at https://osf.io/hxywn/.

Experiment 1

Method

Participants

Prior work using a similar design (Todd et al., 2021, Experiment 2) revealed a small-to-medium sized salience effect on racial bias in the WIT (Salience × Race Prime × Target Object interaction: $\eta_p^2 = .028$). Thus, we set a target sample size ($N = 280$) affording $\geq 80\%$ power to detect $\eta_p^2 = .028$ (Faul et al., 2007). In total, 311 undergraduates consented to participate for course credit. We decided a priori to exclude data from participants who performed at or below chance (errors on $\geq 50\%$ of trials) on any trial type in the WIT ($n = 21$). Retaining the excluded data did not meaningfully alter any of the conclusions in either experiment. The final sample comprised 290 participants (81% women, 15.4% men, 1.8% non-binary; 15% White, 2.1% Black, 57.3% Asian, 17.1% Latino/a/e/x, 5.2% multiracial; $M_{age} = 19.3$, $SD = 1.3$).

Procedure

In both experiments, participants arrived at the lab in small groups and were led by an experimenter to an individual computer workstation to complete the experimental tasks. Participants completed a sequential priming task, the WIT (Payne, 2001), wherein two images appeared in quick succession. Instructions urged participants to ignore the first image (face prime) and to classify the second image (target object) quickly and accurately via key press. The face primes were facial images of 48 men varying in apparent race (24 Black, 24 White) and posed emotion expression (24 angry/scowling, 24 happy/smiling) from the Chicago Face

Database (Ma et al., 2015).[3] The target objects were 6 gun and 6 tool images from Payne (2001). Each trial comprised the following sequence: fixation cross (500 ms), face prime (200 ms), target object (200 ms), and pattern mask (until participants responded). If participants failed to respond within 500 ms, a message ("Please respond faster!") appeared (1 s).

We structured the WIT so that apparent race or emotion expression was more distinctive throughout the task (Macrae & Cloutier, 2009; Rees et al., 2022; Todd et al., 2021). Participants were randomly assigned to complete one of two WIT variants, each comprising two blocks of 144 experimental trials (288 total trials that were preceded by 12 practice trials). In the *race-salient* condition, the face primes were scowling Black and White men in one block of trials and smiling Black and White men in the other block. In the *expression-salient* condition, the face primes were smiling and scowling Black men in one block of trials and smiling and scowling White men in the other block. Within a given block of trials, varying only one source of information (e.g., emotion expression) should render it more contextually distinctive, and thus more salient (Taylor & Fiske, 1978), than the other source of information (e.g., apparent race). Block order was counterbalanced and did not moderate racial bias.

Analysis Plan

Prior to all analyses, we excluded trials with RTs <100 ms and >1500 ms (Todd et al., 2021), which eliminated 2.4% of the data in both experiments. We also excluded error trials prior to RT analyses (but see the Supplementary Materials for RT analyses of error trials). Below, we report analyses pertinent to our focal hypotheses on information salience effects on racial bias.

[3] The emotion expression and apparent race of these face stimuli were likely construed unambiguously. In the face categorization task in Experiment 2, emotion expression and race were both "correctly" classified on ≥ 95% of trials, supporting the assumption that both sources of information were clear and easy to identify (see Tables S7 & S8).

Behavioral Data Analyses. All analyses were conducted using linear mixed-effects models (LMEMs), with each model containing fixed effects for Salience, Race Prime, Emotion Prime, Target Object, and all identifiable interactions. Models included a random-effects structure with by-participant and by-stimulus random intercepts.[4,5] This approach is analogous to fitting the data to a mixed analysis of variance that is adjusted for the cross-classified clustering of responses within participants and within stimuli. We examined our effect of interest (i.e., the Salience × Race Prime × Target Object effect) via contrasts of the model's Race Prime × Target Object interactions across and between salience conditions. Full LMEM tables appear in the Supplementary Materials (see Tables S1 and S3).

DDM Parameter Estimation. For each experiment, we estimated the model using a Markov Chain Monte Carlo (MCMC) sampler in JAGS 4.30 (Plummer, 2003) with the Wiener distribution provided by Wabersich and Vandekerckhove (2014) and an estimation approach to make inferences in this framework (Gelman et al., 2003; Kruschke, 2014). Mirroring model specifications from Todd et al. (2021; see also Pleskac et al., 2018), all parameters were allowed to vary by Race Prime, Emotion Prime, and Salience, and the drift rate and non-decision time parameters also were allowed to vary by Target Object.[6] As in prior work (Todd et al., 2021), posterior predictive checks suggest that the model adequately characterizes the WIT data. The representativeness and accuracy of each model's estimation were assessed both visually and

[4] The only exception was for the LMEM on incorrect response times reported in the Supplementary Materials. Due to boundary fit conditions, we removed the by-stimulus random intercept from the model.
[5] Although the LMEM on error rates in Experiment 2 afforded the inclusion of by-participant random slopes for Race Prime, we chose to prioritize consistency within and across experiments over that single model's random effects structure. Inclusion versus exclusion of the additional random effect did not meaningfully change the results.
[6] As highlighted in Table 1, the threshold separation and relative start point parameters cannot be identified across conditions of Target Object: The relative start point parameter reflects the position at which participants are closer to a gun versus tool decision *at target onset*; the threshold separation parameter reflects the extent to which evidence must be accumulated to reach a gun versus tool decision, presumably determined before target onset. Presumably both the extent of evidence accumulated from the target object (i.e., drift rate) and the processing time prior to a response being recorded (i.e., non-decision time) may vary by Target Object.

numerically (see the Supplementary Materials) and were found to be adequate enough to rely on the parameter estimates for subsequent process analyses.

To compare parameter estimates across conditions, we computed contrasts that included the 95% highest density interval ($HDI_{95\%}$) of the difference between posterior distributions of each parameter across the relevant conditions. Differences with $HDI_{95\%}$ excluding 0 are considered credible. For each analysis, we report the most credible estimate of the raw difference, a Cohen's *d*, and the $HDI_{95\%}$ around *d*. The effect of Race Prime was compared across levels of Salience for all four parameters. For drift rate and non-decision time, contrasts were further computed to evaluate the effect of Race Prime across levels of Target Object. (Figure 2 displays the relative start point parameter estimates in both experiments; Figures S1–S3 display all other parameter estimates).

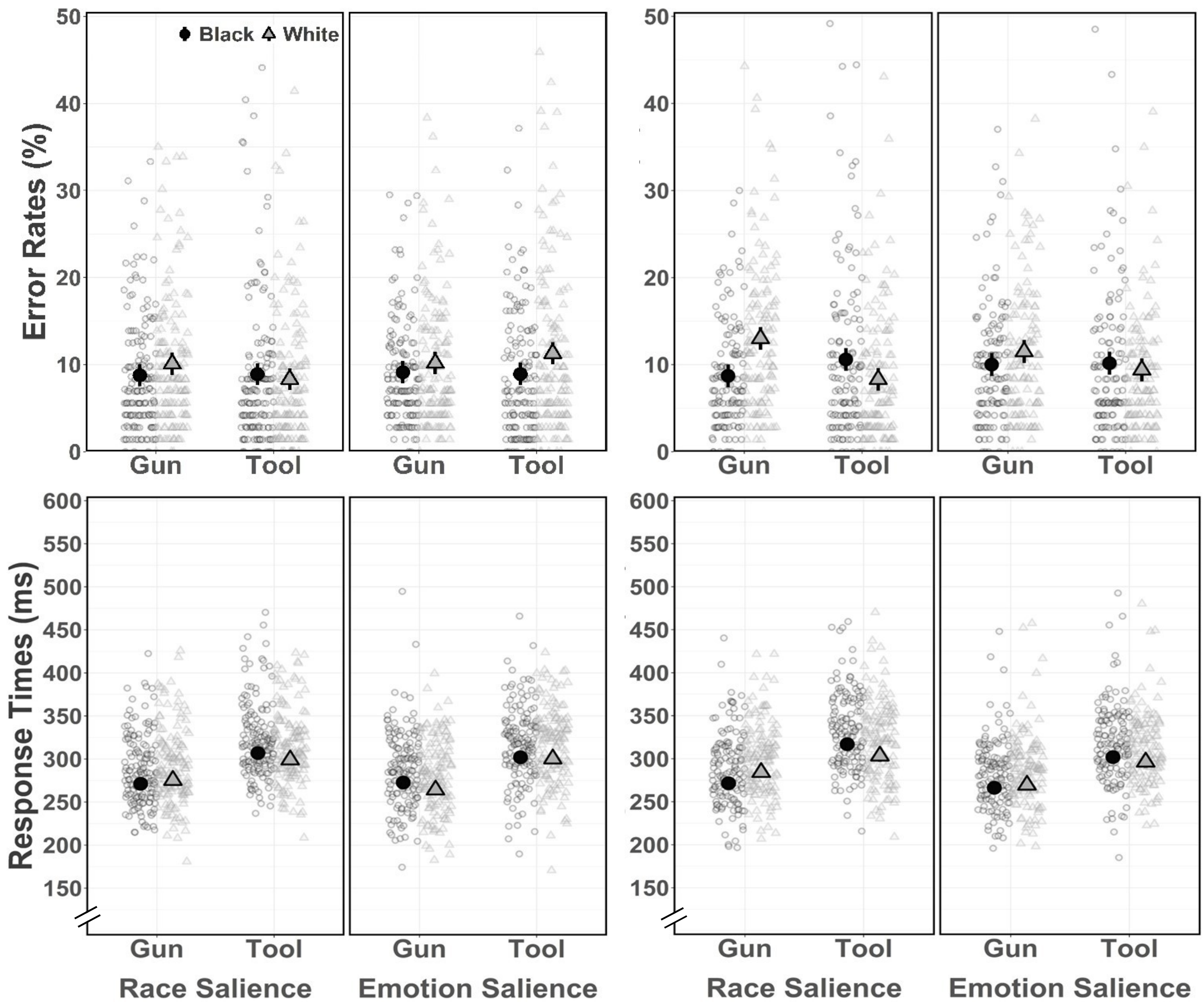
Results

Behavioral Analyses

Error Rates. A significant Salience × Race Prime × Target Object interaction, $\beta = 0.03$, $F(1, 82166.6) = 16.92$, $p < .001$, $R^2 < .01$, revealed salience-driven variation in racially biased weapon identification. When race was salient, the Race Prime × Target Object interaction (i.e., racial bias) was significant, $b = -0.02$, $z = -2.60$, $p = .009$, though neither underlying simple effect of Race Prime reached significance (see Table S5 for simple effects). When emotion was salient, however, the Race Prime × Target Object interaction was not significant, $b = 0.01$, $z = 1.79$, $p = .073$ (see Table S6 for descriptive statistics for each experiment).

Figure 1

Behavioral Data Plots by Race Prime and Salience Condition Across Experiments



Notes. Markers reflect error rates (top row) and correct response times (bottom row) for Black and White prime trials. Empty markers reflect individual-level data and filled shapes and their error bars reflect the estimated marginal means from the linear mixed-effects model applied to those data. The x-axis displays whether the target object was a gun or tool. Shading and shape of markers reflect whether the target object followed a Black or White face prime. Panels vary by salience condition, whereby panels on the left within each plot reflect the race-salient condition and panels on the right within each plot reflect the emotion-salient condition. The plots on the left display data from Experiment 1; the plots on the right display data from Experiment 2.

Correct RTs. A significant Salience × Race Prime × Target Object interaction, β = 0.07, $F(1, 74530.0) = 63.83$, $p < .001$, $R^2 = .03$, again revealed salience-driven variation in racial bias.

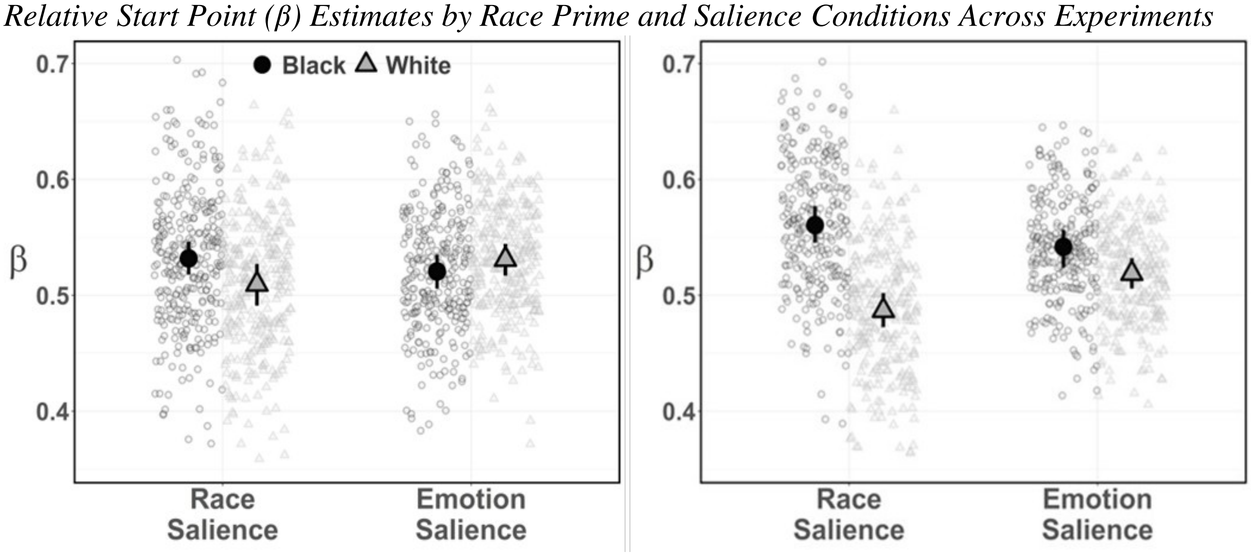
When race was salient, racial bias was evident, $b = -0.04$, $z = -5.11$, $p < .001$. Guns were identified faster ($M_{\text{diff}} = -4$ ms) and tools were identified slower ($M_{\text{diff}} = 6$ ms) after Black versus White primes. Contrary to expectations, when emotion was salient, there was significant racial bias in the opposite direction, $b = 0.03$, $z = 3.16$, $p = .002$. Guns were identified *slower* after Black versus White primes ($M_{\text{diff}} = 7$ ms); the speed of tool identification, by contrast, did not significantly differ between race primes.

Process Analyses

A Salience × Race Prime contrast on the relative start point (β) was credible, $\mu_{\text{diff}} = 0.02$, $d = 0.25$, $\text{HDI}_{95\%}$ [0.11, 0.41]. When race was salient, the decision process began closer to “gun” after Black versus White primes, $\mu_{\text{diff}} = -0.02$, $d = -0.33$, $\text{HDI}_{95\%}$ [-0.54, -0.13]. When emotion was salient, no credible racial bias emerged, $\mu_{\text{diff}} = 0.01$, $d = 0.18$, $\text{HDI}_{95\%}$ [-0.03, 0.38]. These findings align with an initial bias account: Salience-driven variation in racially biased starting positions in the decision process explain salience-driven moderation of racially biased behavior.

Figure 2

Relative Start Point (β) Estimates by Race Prime and Salience Conditions Across Experiments



Notes. Markers reflect posterior estimates for Black prime and White prime trials. Empty markers reflect individual-level estimates and filled shapes and their error bars reflect the most credible values and 95% highest density intervals, respectively, from the DDM modeled to the data. The x-axis displays information salience condition (race, emotion). Shading and shape of markers reflect the salience condition. The plot on the left displays estimates from Experiment 1; the plot on the right displays estimates from Experiment 2.

A small but credible race prime effect emerged on the drift rate (δ), μ_{diff} = -0.14, d = -0.15, $HDI_{95\%}$ [-0.25, -0.06], but it did not vary by information salience, μ_{diff} = -0.05, d = -0.06, $HDI_{95\%}$ [-0.16, 0.03], or target object, μ_{diff} = 0.04, d = 0.04, $HDI_{95\%}$ [-0.06, 0.14]. Accumulated evidence from target objects was stronger after Black versus White primes, regardless of whether emotion or race information was more salient or whether the object was a gun or tool.

The race prime effect on threshold separation (α) was not credible, μ_{diff} = -0.02, d = -0.11, $HDI_{95\%}$ [-0.25, 0.02]. Finally, a small but credible race prime effect emerged on non-decision time (τ), μ_{diff} = -0.004, d = -0.10, $HDI_{95\%}$ [-0.20, -0.03], but it did not vary by information salience, μ_{diff} = -0.001, d = -0.02, $HDI_{95\%}$ [-0.12, 0.05], or target object, μ_{diff} < 0.001, d = 0.01, $HDI_{95\%}$ [-0.08, 0.09].

Discussion

In Experiment 1, racial bias was weaker when emotion versus race was salient. Process analyses failed to support an evidence accumulation account of this effect. Neither target object nor information salience moderated the stronger evidence accumulation occurring after Black versus White primes. Rather, process analyses supported an initial bias account: When race was salient, the decision process began closer to "gun" after Black versus White primes. When emotion was salient, no credible start-point bias emerged. Descriptively, however, start points in the emotion-salient condition were *farther* from "gun" after Black versus White primes, mirroring the atypical pattern of RTs in the emotion-salient condition (e.g., *slower* tool identifications after Black versus White primes). Whether behavior assimilates toward (e.g., typical racial bias) or contrasts from (e.g., atypical racial bias) race stereotypes can vary by context (Bless & Schwarz, 2010), raising questions about whether the atypical pattern in Experiment 1 stems from our blocking design. Experiment 2, therefore, aimed to replicate these results using a different manipulation of information salience.

Experiment 2

Method

Participants

Prior work using a similar design (Todd et al., 2021, Experiment 1) revealed a large effect of information salience on racial bias in the WIT (Salience × Race Prime × Target Object interaction: $\eta_p^2 = .139$); however, because smaller effects are of theoretical interest, we set a target sample size ($N = 258$) affording ≥ 80% power to detect $\eta_p^2 = .03$ (Faul et al., 2007). In total, 278 undergraduates consented to participate for course credit. We decided a priori to exclude data from participants who performed at or below chance (errors on ≥ 50% of trials) on the face categorization task ($n = 1$) or on any trial type in the WIT ($n = 20$). We also excluded data from one participant for whom a computer error caused the WIT to abort early. The final sample comprised 256 participants (73.4% women, 24.2% men, 1.2% non-binary; 12.7% White, 1.9% Black, 61.3% Asian, 15.2% Latino/a/e/x, 4.7% multiracial; $M_{age} = 19.4$, $SD = 2.0$).

Procedure

Participants first completed a face categorization task (Todd et al., 2021) wherein they viewed one of two stimulus sets of facial images, each containing a randomly selected batch of 24 of the 48 facial images from Experiment 1. Both stimulus sets contained equal numbers of male faces varying in apparent race and posed emotion expression. Depending on information salience condition, participants were randomly assigned to classify the faces by *race* (Black vs.

White) or by *emotion expression* (angry vs. happy) via key press. The images appeared one-by-one and remained on screen until participants responded, for a total of 72 trials.

Next, participants completed a WIT that deviated from the WIT in Experiment 1 in two ways. First, the face primes were the other set of 24 facial images not used during the face categorization task. We counterbalanced which stimulus set was used for the face categorization task and the WIT. Using different facial stimuli in the two tasks allowed us to rule out an event coding account (Hommel et al., 2001) whereby memory of specific responses toward specific faces in the face categorization task might affect responses toward those same faces in the WIT. Second, the face prime × target object combinations were fully integrated within a single block of 288 experimental trials that were preceded by 12 practice trials.

Results

Behavioral Analyses

Error Rates. A significant Salience × Race Prime × Target Object interaction, $\beta = 0.04$, $F(1, 74212.8) = 25.35$, $p < .001$, $R^2 < .01$, revealed salience-driven variation in racial bias. Race bias was evident when race was salient, $b = -0.02$, $z = -2.60$, $p = .009$. Guns were misidentified as tools less often after Black versus White primes ($M_{\text{diff}} = -4.3\%$) and tools were misidentified as guns more often after Black versus White primes ($M_{\text{diff}} = 2.2\%$). Racial bias was also evident (albeit more weakly) when emotion was salient, $b = -0.02$, $z = -3.04$, $p = .002$. Guns were misidentified as tools more often after Black versus White primes ($M_{\text{diff}} = -1.5\%$), whereas misidentification of tools did not significantly differ between race primes.

Correct RTs. A significant Salience × Race Prime × Target Object interaction, $\beta = 0.59$, $F(1, 66767.8) = 42.52$, $p < .001$, $R^2 = .04$, again revealed salience-driven variation in racial bias.

Racial bias emerged when race was salient, $b = -0.07$, $z = -11.67$, $p < .001$. Guns were identified faster after Black versus White primes ($M_{diff} = -9$ ms) and tools were identified slower after Black versus White primes ($M_{diff} = 8$ ms). Racial bias also emerged (albeit more weakly) when emotion was salient, $b = -0.03$, $z = -3.98$, $p < .001$. Whereas the speed of gun identification did not significantly differ between race primes, tools were identified slower after Black versus White primes ($M_{diff} = 5$ ms).

Process Analyses

A Salience × Race Prime contrast on the relative start point (β) was credible, $\mu_{diff} = 0.02$, $d = 0.40$, $HDI_{95\%}$ [0.25, 0.57]. When race was salient, the decision process began closer to "gun" after Black versus White primes, $\mu_{diff} = -0.07$, $d = -1.17$, $HDI_{95\%}$ [-1.45, -0.94]. Although start-point bias also emerged when emotion was salient, the effect was weaker, $\mu_{diff} = -0.02$, $d = -0.39$, $HDI_{95\%}$ [-0.60, -0.16]. Like Experiment 1, these findings align with an initial bias account.

A small but credible race prime effect emerged on the drift rate (δ), $\mu_{diff} = -0.13$, $d = -0.17$, $HDI_{95\%}$ [-0.27, -0.06], but it did not vary by information salience, $\mu_{diff} = 0.07$, $d = 0.10$, $HDI_{95\%}$ [-0.01, 0.20], or target object, $\mu_{diff} = 0.07$, $d = 0.10$, $HDI_{95\%}$ [-0.02, 0.20]. Stronger evidence was accumulated for the target objects after Black versus White primes, regardless of whether emotion or race information was more salient or whether the object was a gun or tool.

A small but credible race prime effect also emerged on threshold separation (α), $\mu_{diff} = -0.04$, $d = -0.24$, $HDI_{95\%}$ [-0.38, -0.09], but it did not vary by salience, $\mu_{diff} = 0.01$, $d = 0.07$, $HDI_{95\%}$ [-0.07, 0.22]. The amount of evidence required before responding was greater after Black versus White primes, regardless of information salience. No credible effects emerged on non-decision time (τ).

Discussion

In Experiment 2, facial information salience again moderated racial bias in behavior, and these results again were better explained by an initial bias account. The decision process began closer to “gun” following Black versus White primes, but less so when emotion versus race information was more salient. Once again, stronger evidence accumulation following Black versus White primes did not vary by target object or which information was more salient.

General Discussion

In two experiments, we examined if and how the salience of facial information shapes racially biased weapon identification. We manipulated salience either by augmenting the distinctiveness of emotion or race information during the WIT (Experiment 1) or by augmenting participants’ experience in processing emotion or race information prior to the WIT (Experiment 2). Racial bias in behavior was consistently weaker when the salience of emotion versus race information was highlighted. These findings complement a growing body of evidence suggesting that the salience of facial information other than race (e.g., the age of the face primes) can alter racially biased weapon identification (Jones & Fazio, 2010; Todd et al., 2021; see also Gawronski et al., 2010). Specifically, they suggest that attending to comparatively more dynamic and affect-laden information communicated by facial expressions of emotion can likewise moderate racially biased weapon identification.

Using diffusion modeling, we tested competing cognitive accounts of *how* facial information salience shapes racially biased weapon identification. Our results contradict the evidence accumulation account, which posits that evidence is accumulated from stereotype-congruent (vs. stereotype-irrelevant) target objects more strongly following race primes, and that the salience of information in the face primes shapes this phenomenon. In both experiments, the strength of evidence accumulation did not vary stereotypically (e.g., larger estimates for guns following Black primes) nor by information salience.

Our process-level analyses instead consistently supported an initial bias account, which posits that the weapon identification process begins closer to a race-stereotypic decision after encountering race information in the face primes, and that the salience of facial information shapes the strength of this start-point bias. In both experiments, the decision process began closer to "gun" responses shortly after participants encountered Black versus White face primes. Furthermore, the strength of this effect was shaped by the salience of facial information: Racially biased start points were either eliminated (Experiment 1) or attenuated (Experiment 2) when emotion versus race was salient. Considered alongside previous findings of moderation by age salience (Todd et al., 2021), these results support the initial bias account as a mechanism whereby attending to person information besides race lowers the likelihood of favoring the "gun" response before the object's appearance, relative to attending to race-related information.

Notably, our experiments failed to replicate prior findings that the mere availability of emotion cues in face primes moderates racially biased weapon identification (see Tables S1 & S3). Whereas Kubota and Ito (2014) found that racial bias emerged for scowling but not smiling face primes (see also Raissi & Steele, 2021), here emotion expressions in the face primes failed to moderate racial bias (despite these emotion expressions being easily detected; see footnote 3 and Tables S7 & S8). Furthermore, emotion expressions weakly moderated weapon identification (i.e., the Emotion Prime × Target Object interaction) in Experiment 2, but this effect was not moderated by the salience of emotion. This latter point offers further clarity to the question of *how* emotion versus race salience shapes racially biased weapon identification. If racial bias is weaker in the emotion-salient versus race-salient condition because participants attended more to emotion information in the face primes, then the effect of emotion expression on weapon identification (i.e., the Emotion Prime × Target Object interaction) should be stronger when emotion versus race is salient. That is, if participants are paying more attention to emotion, then the effect of emotion expressions should be more impactful. And yet, we found no evidence that emotion salience moderated the impact of emotion expression on weapon identification.

Our findings suggest that the mere availability of obvious emotion expressions *does not* moderate racially biased weapon identification, but that increasing the salience of emotion expressions *does* moderate racial bias, relative to increasing salience of race. And yet, increasing the salience of emotion expressions failed to moderate the effects of emotion expression on weapon identification. It is unclear, therefore, if attention is simply being drawn *away* from race information without being drawn *toward* emotion information. This pattern of results underscores the importance of directly measuring the processing of prime-related content to clarify when and how salient cues are integrated into object identification. The current instantiation of the DDM is ill-equipped to answer this question because it measures the decision process from target object onset, treating the influence of face primes as a response bias toward or away from "gun" decisions (i.e., a start-point bias).

In reality, the processing of facial information occurs over a time course rich in nuance (Freeman et al., 2020). Such nuance may be needed to understand where attention is directed at prime onset, and how such attention allocation affects later-stage processing. For example, the length of time spent processing information in the primes might flip the direction of their impact on decisions about targets (Klauer et al., 2009). We see value in future research that uses alternative computational approaches (e.g., Diederich & Trueblood, 2018) to capture the processing of face primes more directly. By dynamically measuring the processing of face primes in the WIT, future work may identify the amount of processing time required for various information in the face primes to have maximal impacts on later-stage processing.[7]

Future research should also test the generalizability of the initial bias account across other sources of salient facial information and different social groups. For example, information salience also shapes gender-stereotypic threat impressions (Rees et al., 2022), but it remains unclear where in the decision process these effects emerge. In addition, because we used only male face primes, future research should test whether racially biased weapon identification evoked by Black versus White women (Thiem et al., 2019) is likewise shaped by informational salience (cf. Petsko et al., 2022) and, if so, whether it is best explained by an initial bias account.

Our findings indicate that attending to emotion versus race information can weaken racially biased weapon identification. This phenomenon can be explained by salience-driven changes at the start of the decision process. Racial biases favoring a “gun” response before the object’s onset were weaker when emotion versus race was salient, pointing to a mechanism whereby the salience of person information moderates racially biased decision-making.

[7] We thank an anonymous reviewer for raising this point.

References

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, *15*, 88-93.

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 42, pp. 319-373). San Diego, CA: Elsevier Academic Press.

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*, 219-233.

Diederich, A., & Trueblood, J. S. (2018). A dynamic dual process model of risky decision making. *Psychological Review, 125*, 270-292.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 61, pp. 237-287). Elsevier.

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorisation influence spontaneous evaluations of multiply categorisable objects? *Cognition and Emotion*, *24*, 1008-1025.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York: Guilford Press.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849-878.

Johnson, D. J., Cesario, J., & Pleskac, T. J. (2018). How prior information and police experience impact decisions to shoot. *Journal of Personality and Social Psychology, 115*, 601-623.

Jones, C. R., & Fazio, R. H. (2010). Person categorization and automatic racial stereotyping effects on weapon identification. *Personality and Social Psychology Bulletin*, *36*, 1073-1085.

Klauer, K. C., Teige-Mocigemba, S., & Spruyt, A. (2009). Contrast effects in spontaneous evaluations: A psychophysical account. *Journal of Personality and Social Psychology, 96*, 265-287.

Kubota, J. T., & Ito, T. A. (2014). The role of expression and race in weapons identification. *Emotion, 14*, 1115-1124.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*, 1122-1135.

Macrae, C. N., & Cloutier, J. (2009). A matter of design: Priming context and person perception. *Journal of Experimental Social Psychology*, *45*(4), 1012-1015.

Niedenthal, P. M., & Brauer, M. (2012). Social functionality of human emotion. *Annual Review of Psychology*, *63*, 259-285.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181-192.

Payne, B. K., & Correll, J. (2020). Race, weapons, and the perception of threat. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 62, pp. 1-50). Academic Press.

Petsko, C. D., Rosette, A. S., & Bodenhausen, G. V. (2022). Through the looking glass: A lens-based account of intersectional stereotyping. *Journal of Personality and Social Psychology, 123*, 763-787

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

Raissi, A., & Steele, J. R. (2021). Does emotional expression moderate implicit racial bias? Examining bias following smiling and angry primes. *Social Cognition, 39*, 570-590.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*, 260-281.

Rees, H. R., Sherman, J. W., Klauer, K. C., & Todd, A. R. (2022). On the use of gender categories and emotion categories in threat-based person impressions. *European Journal of Social Psychology, 52*, 597-610.

Rivers, A. M. (2017). The weapons identification task: Recommendations for adequately powered research. *PLOS ONE*, *12*, e0177857.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 249-288). New York, NY: Academic Press.

Thiem, K. C., Neel, R., Simpson, A. J., & Todd, A. R. (2019). Are Black women and girls associated with danger? Implicit racial bias at the intersection of target age and gender. *Personality and Social Psychology Bulletin, 45*, 1427-1439.

Todd, A. R., Johnson, D. J., Lassetter, B., Neel, R., Simpson, A. J., & Cesario, J. (2021). Category salience and racial bias in weapon identification: A diffusion modeling approach. *Journal of Personality and Social Psychology, 120*, 672-693.

Todd, A. R., Thiem, K. C., & Neel, R. (2016). Does seeing faces of young Black boys facilitate the identification of threatening stimuli? *Psychological Science, 27*, 384-393.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*, 455-460.